Contents

Original Papers

Appropriateness of Alerts and Physicians’ Responses With a Medication-Related Clinical Decision Support System: Retrospective Observational Study (e40511)
Hyunjung Park, Minjung Chae, Woohyeon Jeong, Jaeyong Yu, Weon Jung, Hansol Chang, Won Cha. .......................................................... 2

The Factors Contributing to Physicians’ Current Use of and Satisfaction With Electronic Health Records in Kuwait’s Public Health Care: Cross-sectional Questionnaire Study (e36313)
Jawaher Al-Otaibi, Eleni Tolma, Walid Alali, Dari Alhuwail, Syed Aljunid. .......................................................... 12

Successful Integration of EN/ISO 13606–Standardized Extracts From a Patient Mobile App Into an Electronic Health Record: Description of a Methodology (e40344)
Santiago Frid, Maria Fuentes Expósito, Inmaculada Grau-Corral, Clara Amat-Fernandez, Montserrat Muñoz Mateu, Xavier Pastor Duran, Raimundo Lozano-Rubi. .......................................................... 19

Fast Healthcare Interoperability Resources for Inpatient Deterioration Detection With Time-Series Vital Signs: Design and Implementation Study (e42429)
Tzu-Wei Tseng, Chang-Fu Su, Feipei Lai. .......................................................... 31

Coronary Artery Computed Tomography Angiography for Preventing Cardio-Cerebrovascular Disease: Observational Cohort Study Using the Observational Health Data Sciences and Informatics’ Common Data Model (e41503)
Woo Bae, Jihoon Cho, Seok Kim, Borham Kim, Hyunyoung Baek, Wongsun Song, Sooyoun Yoo. .......................................................... 43

A Recurrent Neural Network Model for Predicting Activated Partial Thromboplastin Time After Treatment With Heparin: Retrospective Study (e39187)
Sebastian Boie, Lilian Engelhardt, Nicolas Coenen, Niklas Giesa, Kerstin Rubarth, Mario Merk, Felix Balzer. .......................................................... 56

Identifying Patients With Heart Failure Who Are Susceptible to De Novo Acute Kidney Injury: Machine Learning Approach (e37484)
Caogen Hong, Zhoujian Sun, Yuzhe Hao, Zhanghuiya Dong, Zhaodan Gu, Zhengxing Huang. .......................................................... 69
Original Paper

Appropriateness of Alerts and Physicians’ Responses With a Medication-Related Clinical Decision Support System: Retrospective Observational Study

Hyunjung Park1*, BA; Minjung Kathy Chae1, MD, MSc; Woohyeon Jeong1, BA; Jaeyong Yu1, MS; Weon Jung1, BA; Hansol Chang1,2, MD; Won Chul Cha1,2,3*, MD, PhD

1Department of Digital Health, Samsung Advanced Institute of Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea
2Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
3Digital Innovation Center, Samsung Medical Center, Seoul, Republic of Korea
*these authors contributed equally

Corresponding Author:
Won Chul Cha, MD, PhD
Department of Digital Health
Samsung Advanced Institute of Health Sciences & Technology
Sungkyunkwan University
81 Irwon-ro
Gangnam-gu
Seoul
Republic of Korea
Phone: 82 2 3410 2053
Fax: 82 2 3410 0012
Email: docchaster@gmail.com

Abstract

Background: Alert fatigue is unavoidable when many irrelevant alerts are generated in response to a small number of useful alerts. It is necessary to increase the effectiveness of the clinical decision support system (CDSS) by understanding physicians’ responses.

Objective: This study aimed to understand the CDSS and physicians’ behavior by evaluating the clinical appropriateness of alerts and the corresponding physicians’ responses in a medication-related passive alert system.

Methods: Data on medication-related orders, alerts, and patients’ electronic medical records were analyzed. The analyzed data were generated between August 2019 and June 2020 while the patient was in the emergency department. We evaluated the appropriateness of alerts and physicians’ responses for a subset of 382 alert cases and classified them.

Results: Of the 382 alert cases, only 7.3% (n=28) of the alerts were clinically appropriate. Regarding the appropriateness of the physicians’ responses about the alerts, 92.4% (n=353) were deemed appropriate. In the classification of alerts, only 3.4% (n=13) of alerts were successfully triggered, and 2.1% (n=8) were inappropriate in both alert clinical relevance and physician’s response. In this study, the override rate was 92.9% (n=355).

Conclusions: We evaluated the appropriateness of alerts and physicians’ responses through a detailed medical record review of the medication-related passive alert system. An excessive number of unnecessary alerts are generated, because the algorithm operates as a rule base without reflecting the individual condition of the patient. It is important to maximize the value of the CDSS by comprehending physicians’ responses.

(JMIR Med Inform 2022;10(10):e40511) doi:10.2196/40511

KEYWORDS
clinical decision support system; computerized physician order entry; alert fatigue; health personnel; decision-making support; physician behavior; physician response; alert system
Introduction

Background

Computerized physician order entry (CPOE), linked to a clinical decision support system (CDSS), has become essential in the healthcare system. The main purpose of a CDSS is to improve patient safety and quality of care, and a medication-related CDSS is especially valuable [1,2]. In a medication-related CDSS, the alerting system provides dosing guidance or drug-drug, drug-allergy, and drug-age warnings that help clinicians prescribe correct orders. Early studies on CDSSs prompted substantial anticipation that medication-related CDSSs, such as alerting systems, may prevent adverse events and enhance patient safety [3,4].

Despite the increasing implementation of CDSS alerts, a substantial number of alerts are overridden [5-7]. The alert override rate is high, sometimes up to 96% [5]. Override is often invoked for reasons such as low alert specificity (i.e., a lack of clinical relevance) and inadequate alert content [8,9]. Low alert acceptance was associated with repeated alerts that are inappropriate [6,10]. Excessive alerts that are not clinically relevant could lead to alert fatigue and contribute to alert overrides [11,12].

A common issue connected with the implementation of clinical decision support tools in electronic medical records (EMRs) is alert fatigue [13]. Alert fatigue is the issue in which users of a CDSS that generates an excessive amount of warning messages tend to overlook the majority of these alerts, including those that warn them of potentially clinically relevant errors [2]. A CDSS can fail to enhance patient safety due to alert fatigue. Alert fatigue arises when an excessive number of irrelevant alerts drives users to routinely override them [14].

In the CDSS, 2 types of alerts are usually used. One type of alerts is active or “pop-up” warnings. These alerts require an action from the user for the clinical process to continue, such as clicking a button or stating the overriding reason. The other type of alerts is passive warnings, such as flagging potentially abnormal values. Passive alerts, unlike active alerts, do not interrupt the provider’s workflow; hence, these alerts do not require a response from the user to override the clinical process. Numerous studies have established the issue of alert fatigue with active alerts [10,12,15,16]. Passive alerts may also be a substantial cause of alert fatigue. The true burden of these alerts has rarely been assessed [17].

There is limited research evaluating the appropriateness of overrides with no override reasons in the passive alert system and the alert itself for clinical appropriateness for a patient’s specific condition. To understand the behavior of physicians, previous studies have only evaluated the appropriateness of overrides based on their reasoning [1,18]. In this study, we evaluated the appropriateness of alerts and physicians’ responses in a passive alert system through a patient EMR. We also categorized the alerts assessed by clinical relevance and physicians’ responses. This study may provide insights into the clinical use of medication alerts, whether physicians override them, and what reactions physicians offer when responding to them.

Objective

This study aimed to evaluate the clinical appropriateness of alerts and the corresponding physicians’ responses in a medication-related passive alert system.

Methods

Study Design

This study was a retrospective observational study with stratified sampling according to medication. The analyzed alerts were generated from medication orders between August 2019 and June 2020 in the emergency department (ED). We obtained medication orders, alerts, and patient EMR data from a clinical data warehouse (CDW). In Korea, it is stipulated by law that only physicians can prescribe orders, except in a limited number of cases.

Ethics Approval

This study was approved by the Institutional Review Board of the Samsung Medical Center (IRB 2021-09-115).

Study Setting

This study was conducted in the ED of a tertiary academic medical center in Seoul, Korea. It serves 2 million outpatient visits annually and provides in-hospital service for 1975 beds. The ED has 69 beds and approximately 35 doctors. The annual number of patients visiting the ED ranges from 75,000 to 80,000. The workflow of the ED is uncontrolled and unpredictable [19]. Adverse events following an ED visit were reported less frequently but were more preventable than in other hospital settings [20]. Since the ED has various medication prescription patterns, diverse alerts can be analyzed by checking the patients in the ED.

EMR System and Medication Order (Prescription) System

Our EMR system is a self-developed system implemented in 2016. Data Analytics and Research Window for Integrated Knowledge (DARWIN) is an extensive system that includes CPOE as well as nursing, pharmacy, billing, and research support and even patient portal and web services.

CDSS Design: Passive Alert System

A passive alert system in the medication CDSS was applied to the DARWIN. Although passive alerts with in-line text do not interfere with physicians’ workflow, they may also result in decreased effectiveness of the CDSS alerts [21]. The alert appears before the order is confirmed. A response is not required to allow the prescription. The rule-based database for the CDSS was supplied by the KIMS POC knowledge base (KIMS Co) with weekly updates. The types of alerts were age, allergy, dose, drug-drug interaction (DDI), and renal.

CDW Use

This study was performed using data extracted from the CDW at the study site. The CDW is an integrated storage for clinical data that are updated daily, such as deidentified patient data that are updated daily, such as deidentified patient...
demographic information, diagnosis, prescription, and laboratory results. In the past, researchers had to check the variables required for research individually and process the data accordingly. However, using the CDW, researchers can easily obtain the data automatically, sorted according to the various variables assumed by the researcher. CDW supports the automatic conversion of unstructured data, such as text to standardized data, to make it possible to conduct prospective cohort studies conveniently.

Selection of Alerts
In all, 20 frequently overridden medication alerts were selected. We thought that alerts that are frequently overridden would be less clinically relevant; therefore, we prioritized alerts that are frequently overridden as evaluation targets. DDI types and alerts that are difficult to evaluate for clinical appropriateness were excluded as follows: when there was no specific dose setting information for reduction and when the range of dose adjustment according to the indication and severity was wide. Overridden cases and nonoverridden cases were randomly extracted from 20 frequently overridden medication alerts. The number of cases for each medication alert are shown below.

Definition of Alert Overrides and Appropriateness
Alert overrides occur when physicians do not change orders as suggested by the alert. Our previous study defined an alert override as no change in order when an alert occurred on the log data [22]. In this study, however, alert override means no change in order when an alert occurred or a re-order of the same prescription later. In nonoverridden cases, many physicians prescribed the nonoverridden order again, and we considered this case to be an override. If the identical prescription that generated the alert was given to the same patient within 48 hours, it was deemed an override. Alert clinical relevance means that the alert is suitable for each patient’s condition and that the alert actually helped the physician order the prescription. The physicians’ response appropriateness indicates whether the physicians’ override or nonoverride was appropriate considering the patient’s clinical condition.

Detailed Medical Record Review
Through advanced medical record reviews of alert overridden cases and literature research, a group of 3 clinicians (a physician, a pharmacist, and a nurse) determined the criteria for the appropriateness of each alert. In a detailed medical record review, information such as the patient’s age, gender, weight, laboratory results (potassium, sodium, serum creatinine, or glomerular filtration rate, etc), and computed tomography status was confirmed through the patient’s EMR. Each group member independently reviewed random samples of the 382 alert cases for the evaluation of the appropriateness of alert clinical relevance and physicians’ responses. When panel members disagreed, consensus was reached via group discussion.

Classification of Alerts
The alerts were classified based on the results of the appropriateness evaluation. We referred to the evaluation framework developed by McCoy et al [23]. Since the passive alert system does not collect the overriding reason, it may be difficult to judge the appropriateness. Therefore, we included a nondecidable category in the alert classification table (Figure 1).

Figure 1. Classification table for alerts. The alert classification table included the nondecidable category—since the passive alert system does not include an override reason, some cases might be difficult to evaluated.

**Physicians’ response**

<table>
<thead>
<tr>
<th>Alert clinical relevance</th>
<th>Appropriate</th>
<th>Inappropriate</th>
<th>Nondecidable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriate</td>
<td>Successful alerts</td>
<td>Physicians’ nonadherence</td>
<td>Response nondecidable</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>Justifiable overrides</td>
<td>Unintended adverse consequences</td>
<td>Response nondecidable</td>
</tr>
<tr>
<td>Nondecidable</td>
<td>Alert nondecidable</td>
<td>Alert nondecidable</td>
<td>nondecidable</td>
</tr>
</tbody>
</table>

Korean Triage and Acuity Scale (KTAS)
The KTAS is an evaluation tool used to categorize the severity and urgency of ED patients. It is a 5-level triage scale based on the severity of the patient’s chief complaint and symptoms. The KTAS was established in 2012 in Korea in an effort to enhance patient safety and minimize ED congestion at the hospital level. Patients who enter the ED are evaluated by KTAS using the following procedure: impression evaluation, infection confirmation, primary symptom selection, and primary/secondary considerations [24,25].

Data Analysis
Commonly overridden medications were subgrouped according to alert type, and alert patterns were examined. Samples for the medical record review were extracted using stratified random sampling. In our samples, we analyzed the appropriateness of alerts, physicians’ responses, and patient demographics. Interrater reliability for the evaluation of alert and physicians’
response appropriateness was calculated by using a \( \kappa \) index. The results are presented as counts and percentages. The rate of false positive alerts, physicians’ response inappropriateness, and override were expressed as percentages of total alerts. All statistical tests were performed using R statistical software (version 4.0.3; R Foundation for Statistical Computing).

**Results**

Figure 2 shows the detailed selection process for medication alert data. A total of 39,286 (10.5% alert rate) CDSS alerts occurred for 374,133 medication orders between August 2019 and June 2020. We selected 20 frequently overridden medication alerts stratified by the medication alert type (Table 1). The number of alert cases analyzed for medical record reviews was 382 (200 overridden and 182 nonoverridden cases).

The medical record review included 356 patients. Table 2 shows the demographic information of the patients in the medical record review cases. Overall, the patients’ basic characteristics showed that the majority were men (204/356, 57.3%), aged more than 60 years (205/356, 57.6%), and had KTAS scores of 3 (197/356, 55.3%).

A total of 728 medications triggered an alarm; however, we chose 20 frequently overridden medication alerts, because we thought that alerts that are frequently overridden would be less clinically relevant. Table 1 shows the 20 analyzed medications. In the overridden case, all medication alerts included 10 cases; however, in the nonoverridden case, methylprednisolone (n=6), epinephrine (n=9), cefditoren (n=2), cefazolin (n=6), and ampicillin/sulbactam (n=9) had fewer than 10 cases.

Table 3 shows the results of the appropriateness evaluation for alert clinical relevance and physicians’ responses. Interestingly, of the 382 alert cases, the only 7.3% (n=28) were clinically relevant alerts. In the physicians’ response assessment, 92.4% (n=353) were appropriate and 1.6% (n=6) were nondecidable. The interrater reliability for alert clinical relevance appropriateness and physicians’ response appropriateness were moderate (\( \kappa =0.47 \)) and fair (\( \kappa =0.28 \)), respectively. In our study, there was no difference in the appropriateness of clinical relevance between overridden and nonoverridden alerts. When an overridden alert and a nonoverridden alert were classified using a data log rather than a medical record review, the alert appropriateness was 7% (14/200) for overridden alerts and 7.7% (14/182) for nonoverridden alerts, which did not show clinical relevance. Contrary to the expectation that there were more inappropriate alerts in nonoverridden alerts, there was no difference in alert appropriateness between the 2 types of alerts (Multimedia Appendix 1).

In the classification of the 382 alerts, only 3.4% (n=13) were successfully triggered, and 2.1% (n=8) were inappropriate for both the alert and physicians’ response (Table 4). Only 3.9% (n=15) of alerts represented physicians’ nonadherence, where the alert was appropriate but the corresponding physicians’ response was inappropriate. The override rate was 92.9% (n=355): (Physicians’ nonadherence [n=15] + justifiable overrides [n=340])/total alerts [n=382] (Table 4). There were 6 (1.6%) cases in which the physicians’ response could not be determined.

**Figure 2.** Study flow chart. DDI: drug-drug interaction.
<table>
<thead>
<tr>
<th>Order (medication type)</th>
<th>Alert type</th>
<th>Alert counts, n</th>
<th>Overridden alerts for medical record reviews (N=200), n</th>
<th>Nonoverridden alerts for medical record reviews (N=182), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium bicarbonate, 8.4%, 20 mL (other)</td>
<td>Dose</td>
<td>2125</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Esomeprazole, 40 mg (proton pump inhibitor)</td>
<td>Dose</td>
<td>1885</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Ceftriaxone sodium, 2 g (antibiotic)</td>
<td>Renal</td>
<td>1379</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Kalimate powder, 5 g (other)</td>
<td>Dose</td>
<td>1494</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Tazoferan, 2.25 g (antibiotic)</td>
<td>Renal</td>
<td>1108</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Calcium gluconate, 2 g/20 mL (calcium)</td>
<td>Dose</td>
<td>1230</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Acetaminophen, 1 g/100 mL (analgesic)</td>
<td>Dose</td>
<td>1527</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pantoprazole, 40 mg (proton pump inhibitor)</td>
<td>Dose</td>
<td>1059</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Lactulose syrup (other)</td>
<td>Dose</td>
<td>701</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Propacetamol, 1 g (analgesic)</td>
<td>Age</td>
<td>1205</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Methylprednisolone, 4 mg (steroid)</td>
<td>Dose</td>
<td>378</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Ibuprofen, 20 mg/mL (NSAIDs\textsuperscript{a})</td>
<td>Dose</td>
<td>611</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Levofoxacin, 750 mg (antibiotic)</td>
<td>Renal</td>
<td>421</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Terlipressin acetate, 1 mg (vasoconstrictor)</td>
<td>Dose</td>
<td>386</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Epinephrine, 1 mg (other)</td>
<td>Dose</td>
<td>340</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Amiodarone, 150 mg (antiarrhythmic)</td>
<td>Dose</td>
<td>329</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Meropenem, 500 mg (antibiotic)</td>
<td>Renal</td>
<td>301</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Ampicillin/sulbactam, 1.5 g (antibiotic)</td>
<td>Dose</td>
<td>271</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Cefazolin, 1 g (antibiotic)</td>
<td>Dose</td>
<td>275</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Cefditoren pivoxil, 100 mg (antibiotic)</td>
<td>Dose</td>
<td>301</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

\textsuperscript{a}NSAID: nonsteroidal anti-inflammatory drug.
Table 2. Patient demographic.

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Patient (N=356), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>152 (42.7)</td>
</tr>
<tr>
<td>Male</td>
<td>204 (57.3)</td>
</tr>
<tr>
<td><strong>Age (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>0 to 20</td>
<td>58 (16.3)</td>
</tr>
<tr>
<td>20 to &lt;40</td>
<td>18 (5.1)</td>
</tr>
<tr>
<td>40 to &lt;60</td>
<td>75 (21.1)</td>
</tr>
<tr>
<td>≥60</td>
<td>205 (57.6)</td>
</tr>
<tr>
<td><strong>KTAS(^a) score, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>1 (most critical)</td>
<td>13 (3.7)</td>
</tr>
<tr>
<td>2</td>
<td>51 (14.3)</td>
</tr>
<tr>
<td>3</td>
<td>197 (55.3)</td>
</tr>
<tr>
<td>4</td>
<td>94 (26.4)</td>
</tr>
<tr>
<td>5 (least critical)</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td><strong>Injury, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Noninjury</td>
<td>68 (19.1)</td>
</tr>
<tr>
<td>Injury</td>
<td>288 (80.9)</td>
</tr>
<tr>
<td><strong>Disposition, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Discharge</td>
<td>121 (34)</td>
</tr>
<tr>
<td>Admission</td>
<td></td>
</tr>
<tr>
<td>General ward (n=193)</td>
<td>193 (54.2)</td>
</tr>
<tr>
<td>Intensive care unit (n=193)</td>
<td>165 (85.5)</td>
</tr>
<tr>
<td>Transfer</td>
<td>22 (6.2)</td>
</tr>
<tr>
<td>Death</td>
<td>20 (5.6)</td>
</tr>
</tbody>
</table>

\(^a\)KTAS: Korean Triage Acuity Scale.

Table 3. Appropriateness of alert clinical relevance and physicians’ response.

<table>
<thead>
<tr>
<th>Appropriateness evaluation</th>
<th>Case (N=382), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Appropriate</td>
</tr>
<tr>
<td>Alert clinical relevance</td>
<td>28 (7.3)</td>
</tr>
<tr>
<td>Physicians’ response</td>
<td>353 (92.4)</td>
</tr>
</tbody>
</table>

Table 4. Evaluation of alerts.

<table>
<thead>
<tr>
<th>Alert clinical relevance</th>
<th>Physicians’ response (N=382), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Appropriate</td>
</tr>
<tr>
<td>Appropriate</td>
<td>13 (3.4)(^a)</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>340 (89)(^c)</td>
</tr>
<tr>
<td>Nondecidable</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\(^a\)Successful alerts.  
\(^b\)Physician’s nonadherence.  
\(^c\)The override rate (355/382, 92.9%) was determined by the sum of these 2 values divided by the total number of alerts.  
\(^d\)Unintended adverse consequences.
Discussion

Principal Findings

In this study, we evaluated the appropriateness of the alerts and physicians’ responses to the medication-related passive alert system through a detailed medical record review. We found that only 7.3% of alerts were clinically appropriate, and 6% of alerts resulted in inappropriate responses from physicians. Alert fatigue is inevitable when a large number of irrelevant alerts are generated for a small number of appropriate alerts. There were a few successful alerts where the alert was appropriate and the physician accepted the alert. Physicians’ nonadherence of alerts could be a result of the ambiguous contents of alerts that did not provide helpful information [26]. Additionally, a high number of inappropriate alerts could be a reason for physicians’ nonadherence [27]. Physicians were less likely to accept alerts as the number of alerts increased, especially for repeated alerts [6]. When considering the cases where the response of the physician was inappropriate, the alerts where the alert was appropriate were almost twice as common as the alerts where the alert was inappropriate. This finding can be explained by habitual override due to numerous inappropriate alerts [28]. A small number of alerts were classified as resulting in unintended adverse consequences. In a few cases, the physicians’ response appropriateness could not be determined, because the passive alert system did not collect the override reasons. There were no cases where the appropriateness of the alert could not be determined.

Many studies have identified the appropriateness of override according to the appropriateness of the alert [1,5,15,29,30], but only a few studies have evaluated the response of physicians [31-33]. Duke et al [31] conducted a randomized controlled trial on DDI alert targets to identify medical staff’s adherence according to context-enhanced alerting. Strom et al [32] analyzed the unintended effects of a nearly hard-stop CPOE prescribing alert. Understanding the physicians’ response to the CDSS is of importance; however, due to the difficulty in analyzing the response, many researchers simply evaluate the appropriateness of the override. Therefore, it is necessary to increase the utility of the CDSS by understanding physicians’ responses.

In our previous study, we reported an override rate of 61.9% [22]. However, in this study, we found that the override rate was 92.9%. There are several reasons for this difference. First, in this study, through medical record reviews, it was confirmed that some cases that were previously evaluated as nonoverridden by log data were clinically overridden. The difference between the override rate when simply using log data and the override rate through a medical record review is large, even within the same system. In this study, the patients’ overall prescriptions were analyzed through a detailed patient medical record review, and the definition of “override” was expanded. In the previous study, the classification of overridden and nonoverridden alerts was based only on log data [22]. In this study, however, more override was detected by the medical record review than in the previous study. It was confirmed that a substantial number of cases classified as nonoverridden by log data were actually overridden. We found that many physicians prescribed the same prescription that was considered deleted because of an alert. The prescription was considered an override if it was reissued to the same patient within 48 hours of the alert being issued. Therefore, the override rate might be higher in studies that did not identify the nonoverridden alerts [15,29,34,35]. To calculate the override rate properly, it is necessary to establish a mechanism for systematically determining overrides. A standardized definition of override is needed for a detailed analysis and comparison of CDSSs. Furthermore, in this study, we chose the target alerts as alerts that are frequently overridden, so it could be a reason for the high override rate. Additionally, the change of the knowledge base of the CDSS from Medi-Span (Wolters Kluwer Health) to KIMS POC (KIMS Co) may have affected the override rate.

Further research should investigate techniques for improving alert accuracy by using machine learning (ML) and artificial intelligence (AI), analyze the passive CDSS that has not been extensively studied, and explore the causal relationship between the number of alerts and the physicians’ responses. Multiple alerts with low clinical relevance reduce physicians’ reliance of alerts. Additionally, many unnecessary alerts can lead to alert fatigue and increase the probability of ignoring truly important alerts [2]. It is necessary to improve the clinical relevance of the alert to increase the physician’s alert reliance and optimize the alert. ML and AI could be potential solutions. By introducing ML, the rule-based alert system can be improved, and by introducing AI, alerts can be generated according to the individual condition of the patient [36,37]. Despite the promise of technological approaches to drug safety, the risk of mistake will persist if these systems are not carefully applied and heavy attention is not made to building safer systems of care [2]. These considerations are required to reduce needless alerts, improve their clinical relevance, and increase physicians’ alert reliance by assessing CDSS consistently.

Limitations

Our study had several limitations. First, it was performed at a single center with ED practices. Second, the evaluation of physicians’ response appropriateness may be subjective, because passive alert systems do not collect the override reasons. In addition, we did not confirm the clinical consequences of alerts for unintended adverse consequences. Only the clinical consequences related to the prescription stage were checked, and the dispensing/administration stage was not analyzed.

Conclusions

We evaluated the appropriateness of the alerts and physicians’ responses through a detailed medical record review of the medication-related passive alert system. Only by gaining better knowledge of the physicians’ overall behavior is it possible to improve the effectiveness of the CDSS. In our study, most alerts did not reflect the clinical situation of each patient; however, the physicians’ responses were mostly appropriate. Alert fatigue is unavoidable when a large number of irrelevant alerts are generated in response to a small number of useful alerts. It is necessary to decrease unnecessary alerts, improve their clinical relevance, increase alert reliability, and optimize alerts.
Authors’ Contributions

WCC conceived and designed the experiments; HP performed the experiments; MKC, W Jeong, and HC contributed to the experiments; W Jung and YJ analyzed the data; and WCC and HP wrote the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of alert appropriateness according to overridden and nonoverridden alerts. There was no difference in the appropriateness of clinical relevance between overridden alerts (7% appropriate) and nonoverridden alerts (7.7% appropriate).

References


Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
CDW: clinical data warehouse
CPOE: computerized physician order entry
DARWIN: Data Analytics and Research Window for Integrated Knowledge
DDI: drug-drug interaction
ED: emergency department
EMR: electronic medical record
KTAS: Korean Triage and Acuity Scale
ML: machine learning

©Hyun Jung Park, Min Jung Kathy Chae, Woohyeon Jeong, Jaeyong Yu, Weon Jung, Hansol Chang, Won Chul Cha. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 04.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.
The Factors Contributing to Physicians’ Current Use of and Satisfaction With Electronic Health Records in Kuwait’s Public Health Care: Cross-sectional Questionnaire Study

Jawaher Al-Otaibi1*, MPH; Eleni Tolma2,3*, MPH, PhD; Walid Alali1*, PhD; Dari Alhuwail4*, PhD; Syed Mohamed Aljunid1*, MD, PhD

1Department of Health Policy and Management, College of Public Health, Kuwait University, Kuwait City, Kuwait
2Social Behavioral Sciences, College of Public Health, Kuwait University, Kuwait City, Kuwait
3Department of Education, European University Cyprus, Nicosia city, Cyprus
4College of Life Sciences, Kuwait University, Kuwait City, Kuwait

*all authors contributed equally

Corresponding Author:
Jawaher Al-Otaibi, MPH
Department of Health Policy and Management
College of Public Health
Kuwait University
Sabah Al-Salem University City, Kuwait Ministry of Health Bldg
Medical Record Jamal Abdel Nasser Street, Sulabiakat Club Circle, Al-ASima area
Kuwait City, 12009
Kuwait
Phone: 965 55615073
Email: jawaher309m@gmail.com

Abstract

Background: Electronic health record (EHR) has emerged as a backbone health care organization that aims to integrate health care records and automate clinical workflow. With the adoption of the eHealth care system, health information communication technologies and EHRs are offering significant health care advantages in the form of error reduction, improved communication, and patient satisfaction.

Objective: This study aimed to (1) investigate factors associated with physicians’ EHR adoption status and prevalence of EHRs in Kuwait and (2) identify factors predicting physician satisfaction with EHRs in public hospitals in Kuwait.

Methods: This study was conducted at Kuwait’s public Al-Jahra hospital from May to September 2019, using quantitative research methods. Primary data were gathered via questionnaires distributed among 295 physicians recruited using convenience sampling. Data were analyzed in SPSS using descriptive, bivariate, and multivariate linear regression, adjusted for demographics.

Results: Results of the study revealed that the controlled variable of gender (β=−.197; P=.02) along with explanatory variables, such as training quality (β=.068; P=.005), perception of barriers (β=−.107; P=.04), and effect on physician (β=.521; P<.001) have a significant statistical relationship with physicians’ EHR adoption status. Furthermore, findings also suggested that controlled variables of gender (β=−.193; P=.02), education (β=−.164; P=.03), effect on physician (β=.417; P<.001), and level of ease of use (β=.254; P=.001) are significant predictors of the degree of physician satisfaction with the EHR system.

Conclusions: The findings of this study had significant managerial and practical implications for creating an inductive environment for the acceptance of EHR systems across a broad spectrum of health care system in Kuwait.

(JMIR Med Inform 2022;10(10):e36313) doi:10.2196/36313

KEYWORDS
health informatics; information systems adoption; electronic health record; EHR; public health informatics
**Introduction**

Electronic health record (EHR) systems can provide physicians with accurate information to serve patients more efficiently as compared with paper-based systems [1]. A recent literature review indicated that many health care organizations worldwide, especially in low-income countries, still rely on paper-based systems for maintaining patient records [2]. Research suggests that primary issues faced by traditional systems (ie, paper-based systems) are inaccuracy of information, loss of data, and difficulty in sharing information [3]. In Kuwait, many attempts were made to automate clinical workflows in public hospitals. However, lack of organizational readiness and technical knowledge of the user are primary reasons for EHR implementation failure in Kuwait [4].

Evidence of EHR implementation in public and private health care systems suggests that EHRs are more efficient than paper-based electronic record systems [5]. EHRs significantly improve safety, efficiency, and quality of care provided to patients [6,7].

Furthermore, EHR has a significant impact on the performance of health care workers [6]. An EHR system is an integral part of the clinical decision support system, which provides data to a wide range of health care workers and promptly assists in decisions related to diagnosis and treatment, test results, and the cost of health care [7]. Physicians' efficient use of EHR can decrease medical errors and provide every health care professional with accurate and timely information [8]. Health care workers can access information quickly and efficiently through the EHR system, which aids in diagnoses and follow-up treatments [9,10]. EHR covers various types of information, from patient medical history to assimilated information from laboratories, specialists, pharmacists, and insurance companies. The EHR system is not only confined to inpatient care but also extends to aftercare with local general practitioners [11].

In contrast, electronic medical record (EMR) refers to the electronic chart of a patient’s medical history assessed by the concerned medical staff. Integration of new technologies, such as Internet of Things, machine learning, artificial intelligence, and decision support, into the electronic health care system module and their implementation has transformed health care. Transformation of traditional data center–based solutions into cloud systems have opened new horizons for applying big data, Internet of Things, machine learning, artificial intelligence, and big data analytics to enhance the quality of care offered by health care institutes across the country.

Therefore, this research aimed to identify the EHR system adoption status at Al-Jahra hospital in Kuwait, a public hospital operated by the Ministry of Health offering 1234 beds aided by a surgical suite, emergency services, diagnostic center, and outpatient service. It also aimed to measure physicians’ satisfaction with the EHR system by using the factors influencing their satisfaction. The study’s findings posed significant implications for the public health care system in Kuwait to promote greater use of EHR, which can lead to a decrease in medical errors, better health care services, and overall health care cost reduction [8]. Furthermore, digital transformation is the foremost influential agenda for Kuwait Vision 2035 to strengthen investment in high-quality health care and increase the efficiency of the existing health care system [15].

Recent studies [6,7,9] have focused on the theoretical elements predicting EHR adoption among physicians and satisfaction with the existing EHR system. However, there is limited literature on the theoretical aspects that empirically explain the phenomenon of satisfaction with the use of the EHR system [7]. Therefore, this study mainly aimed to fill the research gap by answering critical questions associated with EHR adoption and the degree of physician satisfaction with the existing EHR system used at Al-Jahra hospital. The study aimed to achieve the following research objectives:

- To investigate the prevalence of EHR and EHR adoption status among physicians at Al-Jahra public hospital.
- To investigate the level of satisfaction with EHR use among physicians working at Al-Jahra hospital.
- To investigate factors predicting physician satisfaction with the EHR system at Al-Jahra hospital.

Kuwait provides high-standard health care coverage to its residents. In governmental facilities, free medical treatment is offered to all Kuwaiti nationals. In contrast, foreign residents must pay an annual fee and nominal charges at every visit to access public health care facilities. Kuwait’s government spends 4.6% of its gross domestic product on public health expenditures. Kuwaits’s health care sectors accounted for 11% of the public spending of Kuwait in 2018. There are currently 97 primary health care centers in Kuwait overseen by the Ministry of Health [14].

The history of EMR in Kuwait dates back to 2000, when the Ministry of Health introduced a national EMR system across the entire primary care facilities and hospitals. Moreover, in 2013, a national eHealth strategy was launched that attempted to consolidate all patient health records into a single health record file managed by Kuwait’s Ministry of Health and the department of Information Systems [11].

Evidence from a study by Alnashmi et al [15] in Kuwait has shown that most physicians in primary health care settings favor from using the EHR system; however, they suggested additional functionality improvements through digital signatures, integration with artificial intelligence, data warehousing, and big data analytics to enhance the quality of care offered by public health care systems.
Methods

Setting
This cross-sectional study occurred at Al-Jahra hospital, a general public hospital in Kuwait. Data were collected from May 2019 to September 2019. The sample size was selected using the Raosoft calculator, which suggests that the sample size for the population of staff members working at Al-Jahra hospital requires 295 responses [16]. The researcher incorporated convenience sampling methods to recruit participants from various stratified groups (i.e., gynecology, general physicians, urologists, orthopedics, ENTs, psychiatry, radiology, pathology, cardiology, and gastroenterology).

Theoretical Framework
The Technology Acceptance Model (TAM), an information system theory based on the acceptance and use of technology by people, was used to develop the conceptual framework. TAM provides information on a technology-based framework for understanding the user’s adoption of technology and preference for using the advanced technologies, particularly in the workplace environment [6]. The theory is based on the following two primary factors: perceived ease of use and perceived usefulness of technology.

The TAM and the Unified Theory of Acceptance and Use of Technology (UTAUT) are two popular theories used in explaining the use of EHRs; UTAUT helps gauge the degree of physician satisfaction with EHR, as satisfaction is an antecedent of repeated behavioral intention [17].

Survey Assessment Tool
The survey tool was designed and refined, followed by the pilot testing procedure. The questions included in the survey were based on the information extracted from the literature review. The survey was pilot tested among 33 physicians who had experience using the EHR for clarity, readability, and feedback. The instrument questionnaire reported overall reliability of all items, with $\alpha=.886$, which suggests that the instrument exhibits high internal consistency.

The final version of the survey consisted of 9 sections and 46 items or questions. They were scored on a 5-point Likert response scale ranging from strongly agree to strongly disagree. The survey was translated into Arabic and then back-translated into English. An expert (DA) also reviewed the survey in Health Information Systems in Kuwait to ensure cultural and contextual fit.

The survey’s psychometric properties were established using the Confirmatory Factor Analysis (CFA) [18]. The results of the CFA showed that all items in each construct were retained, with the exception of 2 items in the scale ‘Perception of Barriers to Using EHR.’ The final instrument consisted of 8 main variables and 35 items. All scales were reliable, with the lowest reliability score of 0.717 (for ‘Perception of Barriers to Using EHR’) and the highest score of 0.897 (for ‘Level of Ease of EHR Function’). There were 2 dependent variables. The first was the physician EHR adoption status, and the second was the degree of physician satisfaction with EHR. The independent variables, as directed by the TAM model, include satisfaction with technical support, preference for using a new EHR system, preference to go back to a paper-based system, perception of barriers to using EHR, the effect of the use of EHR on physician, and level of ease of EHR. The demographics measured in the survey were gender, nationality, age group, education, years of experience, work department, and job title. The quality of the related training was also added as an independent variable based on the theoretical insight of the TAM model, which suggests that training paradigms significantly influence behavioral intentions [19].

Inclusion and Exclusion Criteria
The target population consisted of physicians working at Al-Jahra hospital in Kuwait. Inclusion criteria involved (1) employees of Al-Jahra hospital, (2) physicians, and (3) experience using the EHR system in the hospital, whereas exclusion criteria included (1) former employees of Al-Jahra hospital, (2) administrative staff, (3) nurses, (4) technicians, and (5) physicians working with Al-Jahra on a contractual agreement.

Population and Sampling
According to a previous study, 55% of the physicians in Kuwait are already using an existing adopt EHR system [20]. Considering this adoption rate, a finite population size of 503, a 5% error rate, and a design effect of 1, the required sample consisted of 217 research participants. Assuming a nonresponse rate of 20%, a target sample size of 277 physicians was required for the quantitative study.

Figure 1. Perception of training received by physicians at Al-Jahra hospital. EHR: electronic health record.
Ethics Approval

Ethical approval (2019/1093) was obtained from the Kuwait Ministry of Health Ethical Committee. All research participants signed the informed consent form, which clearly stated the study’s purposes, data use, and participants’ safety (ie, confidentiality and anonymity).

Statistical Analysis

The paper survey was self-administered. The response rate was 95%. Missing values were treated in SPSS using missing values analysis, which suggested that missing values were completely at random, and there was no pattern that resulted in the pairwise deletion of data.

The data were analyzed using the IBM Statistical Package for Social Sciences (version 23; IBM Corp) [21]. Descriptive statistics analysis was also conducted, followed by bivariate analysis. The most common test used in the bivariate analysis was the Pearson correlation analysis. The final statistical analysis used was multiple regression analysis to test the contribution of the independent variables to the dependent variables (ie, current use of EHR and satisfaction with EHRs), adjusted for the demographics. This was done in two steps; first, in model 1, only demographic variables were added to the analysis; then, in model 2, both demographics and the independent variables were added. The alpha level set for this study was .05.

Results

Descriptive Statistics

Of 295 participants, the majority of the participants were male (n=242, 82%) and non-Kuwaitis, (n=259, 88%) from India, Egypt, Asia, Africa, and other parts of the Middle East and North Africa (or MENA) region. Most of the respondents were generally young (n=120, 40.7%), between 30 and 39 years of age, and were experienced physicians (n=100, 33.9% had 5-10 years of work experience). Most of them were registrars (n=88, 29.8%) and gynecologists (n=114, 38.2%), as shown in Table S1 in Multimedia Appendix 1.

In terms of behavioral characteristics, almost 2 of 5 of the respondents (n=124, 42%) reported using the EHR system for more than 5 years. There was a lack of consensus among respondents regarding the quality of related training received; for example, 89 (30.2%) reported receiving low-quality training, whereas another 89 (30.2%) reported receiving high-quality training on the EHR system use.

Bivariate Correlation

Regarding bivariate analysis, the degree of physician satisfaction with the EHR system is strongly correlated with the preference for using the new EHR system (r=0.797) and its effect on physician (r=0.744); it was moderately correlated with satisfaction with technical support (r=0.632) and level of ease of EHR system use (r=0.698), as shown in Tables S2 and S3 in Multimedia Appendix 1.

Multiple Regression Analysis

The first series of multiple regression analyses that included all independent variables in the prediction of the EHR adoption status and adjusted for demographic variables showed that the perception of barriers (β=-.0107; P=.04), the effect of the use of EHR on physician (β=.521; P<.001), and training quality (β=.068; P=.005) are significant predictors of physician EHR adoption status (R²=0.56), as shown in Table S4 in Multimedia Appendix 1.

In the second series of multiple regression analyses that included all independent variables in the prediction of the degree of satisfaction with EHR use and adjusted for demographic variables, findings showed that gender (β=-.1931; P=.02), education (β=-.164; P=.03), effect on physician (β=.417; P<.001), and level of ease of EHR use (β=.254; P<.001) are significant predictors of the degree of physician satisfaction with the EHR system (R²=0.62), as shown in Table S5 in Multimedia Appendix 1.

Discussion

Principal Findings

The study’s primary purpose was to examine the psychosocial factors associated with physicians’ use of EHR and satisfaction with the EHR system at Al-Jahra public hospital in Kuwait. Findings of the study show that the level of EHR adoption status can be predicted with the controlled variable of gender along with explanatory variables, that is, training quality, perception of barriers to using EHR, and effect on the physician. Furthermore, findings also suggested that controlled variables (ie, gender and education) along with explanatory variables (ie, effect on physicians and level of ease of EHR system) significantly influence physician EHR adoption status. The gender of the physician can also play an important role in the use of EHR. In our study, females were more likely than males to use the EHR system and were more satisfied with it, as supported by the literature [22].

The study’s findings validate previous studies [12], which highlight the role of risk and trust relationship in predicting EHR adoption status, as findings revealed that the performance and trust relationship implied by the UTAUT model had no impact on physician intention to use an EHR system. This implies that developers, marketers, and medical professionals should improve and optimize patient communication in the EHR system. Our findings validate previous evidence [21] and also suggest that social factors have a negligible effect on physician intention to adopt EHR system, as physicians are driven by their attitudes, ability to control innovation offered by the EHR system, and holistic benefits offered by the system. Findings also validate the role of training in influencing EHR system adoption status among physicians, as evidence from a study by Dunton [23] suggests that training influences perceived usefulness and perceived ease of use as well as external factors, which significantly enhance physician EHR system adoption status.

In terms of the prediction of EHR use, the most important factor was the effect that the use of EHR had on physicians’ work. This implies that physicians will be more inclined toward using the EHR system if they perceive a beneficial effect of the use of EHR on their work. In addition, the length of use of EHR...
also had a positive contribution to the prediction of EHR use. This is not surprising, since using the EHR system for an extended period will lead to adopting the EHR system, according to a study by Liang et al [20].

Regarding the prediction of physicians’ satisfaction with the use of EHR, the most significant contributor was the effect of EHR use on physicians’ work, as supported by the findings of a previous study [24]. Specifically, it was found that the higher the perceptions of the positive effects of EHR on physicians’ work, the more likely it will be for the physicians to be satisfied with the use of the EHR system. The second most important contributor was the degree of ease of EHR use. Consistent with the findings of other studies, as physicians start to experience the ease of using the EHR system, they will start adopting the EHR system [3]. Moreover, another study [25] found that perceived usefulness and perceived ease of use increase the acceptance of using the EHR system and hence the satisfaction with it. As it was explained, accepting the use of the EHR system was an indication of the level of satisfaction of the physicians. Therefore, it was not surprising that this study found the degree of ease of using EHR as an important factor in satisfaction with it. In other words, as physicians perceived the EHR system to be easy to use, they were more likely to use it and experience higher satisfaction levels.

Another unique finding concerning the satisfaction with EHR was related to the academic background of participants. According to the results, this characteristic was a significant factor. This implies that the knowledge and academic experience of the physicians might have an impact on their satisfaction level. Those physicians with higher qualifications will tend to be more satisfied with the EHR system compared with others because they might believe that using the EHR system would allow them to serve the patients better. Evidence from a study [26] demonstrated that physicians with higher levels of education had higher levels of satisfaction with EHR use.

Finally, age was also significant in the prediction of EHR use but in a negative way. Older physicians were more reluctant to use the EHR system, as supported by a previous study [17]. In our study, the demographics were treated as controlled variables. Therefore, their effect on the regression model and other variables was neglected.

From a programmatic perspective, the following are some recommendations for public health professionals in their effort to promote the use of EHR and increase satisfaction with EHR among those who are already using it in governmental hospitals of Kuwait:

- Professionals should first conduct a needs assessment, identify perceived barriers among physicians, and try to address those barriers.
- Public health professionals should focus on improving the functionality of the EHR system and make it as easy as possible to operate; this will encourage physicians to use it more often and rely on the EHR system when seeing patients.
- Public health professionals are advised to emphasize promoting the EHR system’s positive effects on physicians’ work, which could be done through health communication campaigns.

Limitations
There were some limitations to the study. The results of this study were only limited to Al-Jahra hospital, where the study took place. However, this study can be generalized, and the outcomes can be easily applied to other public sector hospitals in Kuwait, as the research examined satisfaction with EHR and adoption of the EHR system. Second, the TAM was used to assess the adoption of EHR and satisfaction with it. However, the TAM and UTAUT models were not fully applied to this study, as the attitude was not examined. Due to the questionnaire being long, the decision was made to shorten it, so more physicians were interested in filling out the questionnaire and participating in the study. The study did not use purposive sampling as convenience sampling for data collection. The use of a random sample might have produced a different result. The findings of the study are not generalizable to all hospitals in Kuwait. Thus, more studies need to be conducted to validate whether other public hospitals exhibit the same phenomenon.

Future Research
Since this study could not cover all aspects that might be useful in examining the satisfaction with EHR system and current use of it, the following future studies must be carried out. First, the theoretical framework should be expanded to include physicians’ attitudes toward using EHR. Research should be conducted that fully uses the TAM by including physicians’ attitudes. It might prove important to examine physicians’ attitudes, since some physicians might have a positive predisposition toward using EHR but still not use it. It would be interesting from a theoretical and programmatic perspective to examine how attitude relates to intention by itself. Second, a follow-up qualitative study through several interviews with senior physicians and hospital officials should be conducted. Such a study will help identify more in-depth information behind using or not using the EHR system. For instance, qualitative research can complement quantitative research results and help us discover the perceived barriers to adopting the EHR system. In addition, qualitative research can help us find answers to surprising results, such as the fact that women are more likely to adopt an EHR system. Research should be conducted that prove important to examine physicians’ attitudes toward using EHR. Research should be conducted that fully uses the TAM by including physicians’ attitudes. It might prove important to examine physicians’ attitudes, since some physicians might have a positive predisposition toward using EHR but still not use it. It would be interesting from a theoretical and programmatic perspective to examine how attitude relates to intention by itself. Second, a follow-up qualitative study through several interviews with senior physicians and hospital officials should be conducted. Such a study will help identify more in-depth information behind using or not using the EHR system. For instance, qualitative research can complement quantitative research results and help us discover the perceived barriers to adopting the EHR system. In addition, qualitative research can help us find answers to surprising results, such as the fact that women are more likely to adopt an EHR system. This study can be replicated in other governmental hospitals in Kuwait to reach a better understanding of how prevalent the use of EHR is and the degree of satisfaction with its use.

Conclusions
There are important takeaways from the results of this study. First, there is still a need to further expand the EHR system adoption at Al-Jahra hospital, since almost 1 in 5 physicians has never used EHR or has used EHR for less than a year. This could be justified as they may have joined the hospital recently. Second, to increase the adoption rate and satisfaction with the current use of EHR among physicians, public health professionals can make the benefits of EHR adoption more visible to the physicians, remove perceived barriers, make the use of the EHR system as easy as possible, and incorporate a high-quality related training, while providing continuous technical support. Results from this study can be helpful to other
governmental hospitals in Kuwait in their efforts to enhance the levels of adoption and satisfaction with the EHR system. The EHR system has many benefits, and it can be fully realized only when all physicians in governmental hospitals in Kuwait fully adopt it.

Acknowledgments
The authors would like to thank Kuwait University, the Ministry of Health, and Al-Jahra Hospital for facilitating this research. This study did not receive any funding.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Tables S1-S5.

References
8. Agbese FAO, Ikonne SN. Electronic health records (EHRs) use and quality healthcare delivery by physicians in tertiary hospitals in federal capital territory, Nigeria. LPP (e-journal) 2021:1-12 [FREE Full text]


Abbreviations

EHR: electronic health record
EMR: electronic medical record
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology

©Jawaher Al-Otaibi, Eleni Tolma, Walid Alali, Dari Alhuwail, Syed Mohamed Aljunid. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 07.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.
Successful Integration of EN/ISO 13606–Standardized Extracts From a Patient Mobile App Into an Electronic Health Record: Description of a Methodology

Santiago Frid1,2, MSc, MD; Maria Angeles Fuentes Expósito1, MSc; Inmaculada Grau-Corral3,4, PhD; Clara Amat-Fernandez5, MSc; Montserrat Muñoz Mateu2,6, MD, PhD; Xavier Pastor Duran1,2, MD, PhD; Raimundo Lozano-Rubi1,2, MD, PhD

1Medical Informatics Unit, Hospital Clínic de Barcelona, Barcelona, Spain
2Universitat de Barcelona, Barcelona, Spain
3FundiSYS, Barcelona, Spain
4mHealth Observatory, Hospital Clínic de Barcelona, Barcelona, Spain
5Fundació Clínic per a la Recerca Biomèdica, Barcelona, Spain
6Oncology Unit, Hospital Clínic de Barcelona, Barcelona, Spain

Corresponding Author:
Santiago Frid, MSc, MD
Medical Informatics Unit
Hospital Clínic de Barcelona
Villarroel 170
Barcelona, 08036
Spain
Phone: 34 932 27 54 00 ext 3344
Email: santifrid@gmail.com

Abstract

Background: There is an increasing need to integrate patient-generated health data (PGHD) into health information systems (HISs). The use of health information standards based on the dual model allows the achievement of semantic interoperability among systems. Although there is evidence in the use of the Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources (SMART on FHIR) framework for standardized communication between mobile apps and electronic health records (EHRs), the use of European Norm/International Organization for Standardization (EN/ISO) 13606 has not been explored yet, despite some advantages over FHIR in terms of modeling and formalization of clinical knowledge, as well as flexibility in the creation of new concepts.

Objective: This study aims to design and implement a methodology based on the dual-model paradigm to communicate clinical information between a patient mobile app (Xemio Research) and an institutional ontology-based clinical repository (OntoCR) without loss of meaning.

Methods: This paper is framed within Artificial intelligence Supporting CAnCer Patients across Europe (ASCAPE), a project that aims to use artificial intelligence (AI)/machine learning (ML) mechanisms to support cancer patients’ health status and quality of life (QoL). First, the variables “side effect” and “daily steps” were defined and represented with EN/ISO 13606 archetypes. Next, ontologies that model archetyped concepts and map them to the standard were created and uploaded to OntoCR, where they were ready to receive instantiated patient data. Xemio Research used a conversion module in the ASCAPE Local Edge to transform data entered into the app to create EN/ISO 13606 extracts, which were sent to an Application Programming Interface (API) in OntoCR that maps each element in the normalized XML files to its corresponding location in the ontology. This way, instantiated data of patients are stored in the clinical repository.

Results: Between December 22, 2020, and April 4, 2022, 1100 extracts of 47 patients were successfully communicated (234/1100, 21.3%, extracts of side effects and 866/1100, 78.7%, extracts of daily activity). Furthermore, the creation of EN/ISO 13606–standardized archetypes allows the reuse of clinical information regarding daily activity and side effects, while with the creation of ontologies, we extended the knowledge representation of our clinical repository.
Conclusions: Health information interoperability is one of the requirements for continuity of health care. The dual model allows the separation of knowledge and information in HISs. EN/ISO 13606 was chosen for this project because of the operational mechanisms it offers for data exchange, as well as its flexibility for modeling knowledge and creating new concepts. To the best of our knowledge, this is the first experience reported in the literature of effective communication of EN/ISO 13606 EHR extracts between a patient mobile app and an institutional clinical repository using a scalable standard-agnostic methodology that can be applied to other projects, data sources, and institutions.

**(JMIR Med Inform 2022;10(10):e40344)** doi:10.2196/40344

**KEYWORDS**

health information interoperability; mobile app; health information standards; artificial intelligence; electronic health records; machine learning

**Introduction**

**Importance of Patient-Generated Health Data**

Traditionally, physicians were the only actor who registered patient data in health information systems (HISs). In recent years, the focus has shifted toward more active participation by patients in their own health care, particularly by means of patient-generated health data (PGHD) [1].

One relevant source of PGHD are wearable devices that connect to the body surface of patients and can transmit data regarding many biological variables. The number of such devices that generate valuable data is growing considerably.

Furthermore, patient experience has been progressively incorporated into health care processes with the objective to optimize them. One of the most relevant measures of outcomes is the patient-reported outcome measures (PROMs), which record patients’ perception of disease, including relevant symptoms and emotional distress [2]. In the context of the increasingly adopted value-based health care model, Michael Porter developed a formula: value = (results that matter to the patient)/costs [3,4]. In this model, it is key that patients report the results that matter most to them using indicators provided by PROMs [5].

Increasingly, all of these data come from patient mobile apps, and they need to be integrated into HISs for their use in the caregiving process (primary use) or for research purposes (secondary use). However, given the large number of HISs that coexist within a single health organization, this proves to be highly challenging.

**Interoperability in Health Information Systems**

To share clinical information in such a way that it can be unequivocally interpreted, both syntactically and semantically, by 2 or more systems, a common health information standard must be used.

European Norm/International Organization for Standardization (EN/ISO) 13606 is a health information standard that seeks to define a rigorous and stable architecture for communicating health records of a single patient, preserving the original clinical meaning. It is based on a dual model proposed by OpenEHR [6] that includes a reference model (with the necessary components, and their constraints, to represent electronic health record [EHR] extracts) and an archetype model (for the formalization of the clinical domain concepts according to the reference model) [7,8]. Thus, EN/ISO 13606 was designed for the exchange of EHR extracts with full meaning and a high compatibility with OpenEHR [9].

The Fast Healthcare Interoperability Resources (FHIR) standard was developed by Health Level 7 (HL7) with the intention to use modern communication standards for the agile creation of health data communication infrastructures [10]. FHIR’s 80/20 rule (focus on 20% of the requirements that satisfy 80% of the interoperability needs) centers on simplicity rather than completeness. FHIR also provides a health information standard to Substitutable Medical Applications and Reusable Technologies (SMART), a framework that enables medical apps to be written once and run unmodified across different health care information technology (IT) systems [11].

EN/ISO 13606’s advantages over FHIR in terms of modeling and formalization of clinical knowledge, as well as flexibility in the creation of new concepts, suggest it could play a role in the communication of EHR extracts with mobile apps, despite the limited existing evidence. This could be particularly useful in complex scenarios of health data exchange between nodes [12].

**The ASCAPE Project**

This paper is framed within the Artificial intelligence Supporting CANcer Patients across Europe (ASCAPE) project, where breast and prostate cancer, 2 of the most prevalent types of cancer, are considered [13]. One of the main purposes of the project is to use powerful artificial intelligence (AI)/machine learning (ML) mechanisms to support cancer patients’ health status and quality of life (QoL) in 4 different pilots [14,15].

Within the ASCAPE project, clinical partners identified previously validated questionnaires used to capture different QoL issues for both types of cancer. AI-based models ingest data from such questionnaires, as well as data regarding daily activity, side effects, and physicians’ interventions, to predict and suggest improvements in patient QoL issues. Hence, ASCAPE prospectively investigates an AI-based approach toward a personalized follow-up strategy for cancer patients focusing on their QoL issues.

The approach chosen in the project to properly process sensitive medical data is federated learning (FL), a decentralized ML technique where local data are used to train shared global models with a central server, keeping the sensitive data locally.
Objectives
The aim of this study was to design and implement a methodology based on the dual model paradigm in order to communicate clinical information between a patient mobile app and an institutional clinical repository, without loss of meaning. This implies a series of specific objectives:

- To conceptually represent information regarding daily activity and side effects by means of ontologies
- To define a set of archetypes based on EN/ISO 13606 for the standardization and consolidation of patient data in clinical repositories
- To create a scalable conversion module for mobile apps, within the Hospital Clínic de Barcelona’s (HCB) environment, to transform local data and generate EN/ISO 13606–compliant EHR extracts
- To validate the methodology through the successful generation and integration of EHR extracts sent from Xemio Research, a patient mobile app, into the institutional ontology-based clinical repository, OntoCR.

Methods
Ethical Considerations
This study was approved by the Hospital Clínic de Barcelona Ethics Committee for Investigation with Drugs (HCB/2020/0971).

Systems and Servers
OntoCR
Traditionally, HISs were developed with a focus on financial and administrative activities, whereas clinical data have been merely translated from paper records to electronic databases. Clinical concepts and the relationships between them have been poorly developed.

OntoCR is an ontology-driven clinical repository conforming to the EN/ISO 13606 standard that uses ontologies for different purposes [16,17]. On the one hand, they define a conceptual architecture centered on the representation of the clinical process and clinical knowledge. By representing a metamodel of health information standards, classifications, and terminologies, OntoCR can also achieve syntactic and semantic interoperability between different HISs. On the other hand, OntoCR uses an ontology that defines the available elements that can be used to build an app. These elements are used by portlets to create a graphical user interface (GUI) deployed in Liferay [18], thus allowing users to access, visualize, enter, and modify structured data through a web-based clinical workstation. OntoCR is linked to the HCB’s EHRs (SAP) using the patient ID, and it can be accessed via SAP or its own website.

Xemio Research
Xemio Research was developed for breast cancer patients, providing them with proper information, allowing the tracking of symptoms, and collecting physical activity data from its users on a daily basis (steps, time of activity, and calories). The deployment of Xemio Research’s backend takes place within the gated area of the HCB, with a dedicated server (CentOS Linux) whose database is modeled object-oriented in PostgresDB without normalized codes for secondary effect or activity references, just literals names in Spanish.

Xemio Research is published in Apple App Store and Google Play Store, with access restricted to study participants. The app was installed on the patient’s phone by the field researcher during the first visit, where the patient provided signed informed consent. This generated a Xemio Research ID, which was then registered and linked to the ASCAPE ID in OntoCR by the field researcher.

ASCAPE Local Edge
Due to the sensitive nature of real patient data and the security and data treatment requirements of the project, the ASCAPE architecture was implemented in a dedicated server (ASCAPE Local Edge) within the HCB’s environment, supervised by the local IT department (Figure 1).

This architecture was deployed using Kubernetes (k8s) [19], an open source software that accelerates the implementation and administration of containers on a large scale. These containers maintain the microservices needed for the functioning of the project; the processes of data extraction, transformation, and load (ETL); the normalization of retrospective data provided by the HCB; patient anonymization; and predictions offered by AI models. The aforementioned normalization of local data is performed by identifying variables of interest and transforming them to the ASCAPE Common Data Model and HL7 FHIR [15], thus generating a uniform ASCAPE-standardized database for training data sets to feed the AI engines. Furthermore, Local Edge generates and updates ASCAPE’s AI predictive models [14], which are shared and evaluated in its accuracy in the federated node.
Figure 1. Information systems within the ASCAPE project. Patients register side effects in Xemio Research, which also tracks patients’ daily steps. These data are standardized using a conversion module within the HCB environment (see the Methodology section, step 3), and it is both stored in the OntoCR and sent to ASCAPE Local Edge, which generates and updates ASCAPE’s AI predictive models, which are shared and evaluated in its accuracy in the federated node. AI: artificial intelligence; ASCAPE: Artificial intelligence Supporting CAncer Patients across Europe; HCB: Hospital Clínic de Barcelona.

Methodology
The methodology comprises a series of steps to achieve successful sharing of standardized clinical information between a patient mobile app and an institutional clinical repository.

Step 1: Definition of Variables to Communicate and Creation of EN/ISO 13606 Archetypes
The first step in the methodology is to define clinical variables that need to be communicated through EHR extracts. Since this study was framed within the ASCAPE project, we identified variables that needed to be registered and could be recorded with Xemio Research:

- Daily activity: date, steps, calories, and duration
- Side effects: date, finding, value, and severity

To share information standardized with EN/ISO 13606, archetypes that define the chosen variables must be created. EN/ISO 13606’s reference model has multiple components, including the entry (“a result of one clinical action, one observation, one clinical interpretation, or one intention”) and its elements (“The leaf node of the EHR hierarchy, containing a single data value”).

Figure 2 shows a mindmap created with the LinkEHR tool [20] of the “side effect” entry archetype. Data types used are those established by the reference model.
Figure 2. Mindmap of the “side effect” archetype (in Spanish), edited with LinkEHR. The “side effect” entry has 4 elements: date, finding, value, and severity.

Step 2: Creation of Ontologies

Once the archetypes are generated, the clinical concepts defined by them must be represented in both systems (mobile app and clinical repository). The functionalities needed to record these variables had already been developed in Xemio Research. For OntoCR, the Medical Informatics Unit at the HCB created the corresponding ontologies to represent these concepts.

A locally developed ontology named Ontoclinic already had a representation of most of the clinical findings that would be used for this project. Hence, the remaining concepts were modeled and added to Ontoclinic, which was later imported into the ASCAPE ontologies. Ontoclinic also includes metaclasses that represent standard classifications and terminologies. Thus, by indicating that a given class is an instance of the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) metaclass, it allows the normalization of concepts (see Figure 3). Both finding and severity were coded with the international edition of SNOMED CT using this approach.

Afterward, both Xemio Research and OntoCR had to model local concepts following the standard. In the first case, this was performed by a conversion module in Local Edge, independent from the app. This component is configured by a text document in JSON format that contains the SNOMED CT codes for each side effect and its severity. The procedure was developed in Python, and it transforms, conceptualizes, and generates daily EN/ISO 13606 EHR extracts with the data of Xemio Research users.

In OntoCR, the modeling was performed by means of ontologies. The HCB Medical Informatics Unit created an ontology that incorporates both EN/ISO 13606 reference and archetype models, enabling the capability of representing clinical data that conform to the standard. Therefore, new ontologies of each entry were created, where the concepts defined in the archetypes were mapped to the EN/ISO 13606 structure.

Figure 3 shows the ontological modeling of concepts described in steps 1 and 2. The upper-left image displays the Secondary_effect class of the Ontoclinic ontology, with its properties date, severity, finding, and value. The lower-left image shows the modeling of the Ontoclinic Severe class with SNOMED CT, which was performed by making the concept an instance of the SCT metaclass, thus allowing its binding to a code Uniform Resource Identifier (URI) and a concept ID. Finally, the right image displays the Secondary_effect class modeled with EN/ISO 13606 as a subclass of EN/ISO 13606 ENTRY, therefore inheriting properties of its superclass. Once the ontologies that represent the clinical concepts are created, they are uploaded to OntoCR (Figure 4), where they will be ready to receive instantiated patient data.
**Step 3: Communication of Standardized Extracts**

After the variables were defined, represented, and standardized in both systems, extracts were ready to be communicated. Xemio Research has integrated services that transmit extracts with pseudo-anonymized data of either side effects or daily activity collected by the app to an Application Programming Interface (API) in OntoCR, which allows the insertion of extracts into the ontology. This way, instantiated data of patients are stored in OntoCR.

Regarding data security and privacy, Xemio Research generated extracts with anonymous identifiers that were assigned to the patients during recruitment. OntoCR stores the information of both Xemio Research IDs and ASCAPE IDs, so it can integrate the data from the extracts with the rest of the clinical records. Therefore, there is no need for the app to receive data from the hospital’s HIS, which is why communication between Xemio Research and OntoCR is unidirectional. This ensures the confidentiality of the real patient data that are managed.

An example of an EN/ISO 13606 EHR extract of side effects is displayed in Figure 5, where the “Wakefulness” finding (coded with the SNOMED CT concept ID 365930002) is recorded.

**Figure 5.** Example of a deidentified EHR extract of side effects. EHR: electronic health record.
Patients enter information on Xemio Research, which normalizes it through a conversion module, thus creating EN/ISO 13606 EHR extracts. These extracts are sent to the API of OntoCR, which inserts patient data into the ontology. The lower image displays a list of instances of side effects, with the corresponding values of the properties date, value, severity, and finding entered in Xemio Research by the patient. Furthermore, an instance of the re_id EN/ISO 13606 property was inserted, indicating the unique identifier by which this instance is referenced in the EHR system.

The process for developing the communication of extracts started in November 2020 and finished in March 2022, with effective deployment in a production environment. On March 14, 2022, all EHR extracts corresponding to retrospective data were sent, and thereafter, extracts were sent daily.

**Figure 6.** Overview of the process of knowledge modeling and extract communication and integration into OntoCR. Blue arrows indicate knowledge-related processes, while red arrows indicate data-related processes. API: Application Programming Interface; EHR: electronic health record; ISO: International Organization for Standardization; SNOMED CT: Systematized Nomenclature of Medicine – Clinical Terms.
Results

EN/ISO 13606 EHR Extracts
We achieved effective communication of EN/ISO 13606–standardized EHR extracts between a mobile app for patients, Xemio Research, and an institutional clinical repository, OntoCR.

In our study pilot, 62 patients were allocated to use Xemio Research. There were 12 (19.4%) dropouts: 7 (58%) due to a lack of response to questionnaires, 2 (17%) due to medical issues, 2 (17%) lost to follow-up, and 1 (8%) for personal reasons. Furthermore, 3 (4.8%) patients never used the app, leading to a total of 47 (75.8%) users.

Table 1 shows the number of each type of extracts exchanged between December 22, 2020, and April 4, 2022, and the number of patients they pertain to.

When comparing the extracts to the data registered in both Xemio Research and OntoCR databases, no missing or unclear data were detected in the process for the study cohort.

Table 1. Number of extracts communicated throughout the study.

<table>
<thead>
<tr>
<th>EHR archetype</th>
<th>Extracts (N=1100), n (%)</th>
<th>Patients (N=47), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side effects</td>
<td>234 (21.3)</td>
<td>34 (72.3)</td>
</tr>
<tr>
<td>Daily activity</td>
<td>866 (78.7)</td>
<td>38 (80.9)</td>
</tr>
</tbody>
</table>

*EHR: electronic health record.

Archetypes and Ontologies

Furthermore, the methodology created for this project resulted in a series of deliverables within each step of the process. First, the creation of EN/ISO 13606–standardized archetypes allows the reuse of clinical information for the variables considered in this study: daily activity (date, steps, calories, and duration) and side effects (date, finding, value, and severity).

In addition, by creating ontologies that represent the aforementioned clinical variables and integrating them into OntoCR, we continue to extend the knowledge representation of our ontology-based clinical repository.

Discussion

Principal Findings

We describe a methodology for communicating EN/ISO 13606 EHR extracts between a patient mobile app and an ontology-based clinical repository. Standardized information regarding side effects or daily activity of patients enrolled into Xemio Research in the study was effectively communicated.

EN/ISO 13606 was chosen for this project because of the operational mechanisms it offers for data exchange and its advantages regarding modeling of clinical knowledge and flexibility in the creation of new concepts, which is also why it was used in the first place to extend OntoCR’s metamodel with the incorporation of the reference and archetype models of the standard. However, due to the flexibility and standard-agnostic nature of our methodology, there is complete independence regarding any specific standard. Thus, we are able to carry out transformations between health information standards with minimum use of resources and without the need for changes in the database structure.

LinkEHR offers the possibility to create clinical information models using multiple health information standards (EN/ISO 13606, OpenEHR, FHIR) as well as terminologies and classifications (SNOMED CT, International Classification of Diseases 10th Revision [ICD-10], Logical Observation Identifiers Names and Codes [LOINC]), all of which can also be incorporated into OntoCR by creating corresponding metamodel ontologies. The API that inserts instantiated patient data into the repository is prepared to receive any EN/ISO 13606 EHR extract, and it can be extended to incorporate other standards as well. All this facilitates the application of our methodology to other projects and institutions.

Single vs Dual Models and Semantic Interoperability in Health Care

Health information interoperability is one of the requirements for the continuity of health care [21]. The dual model allows the separation of knowledge and information in EHR systems, with the consequent possibility of extending the concept model without the need for specific developments and introducing new concepts when the system is already implemented [22]. With the use of formal information models built from common components and linked to standard terminologies [23], 2 systems can achieve semantic interoperability without prior agreement [24,25].

Single, nonstandardized models require the development of specific interfaces to communicate information with other systems. In a context where there is a growing number of information systems within each health organization, many of which come from mobile devices of both patients and physicians, the scalability of this approach is considerably reduced. These difficulties are even greater when considering the communication of health information between different organizations.

The benefits of standardizing EHR data are not limited to primary use. The reuse of clinical data for secondary purposes, such as investigation in both single- and multicenter studies, requires formal information models in order to make data unequivocally understandable and reproducible [26].

Comparison With Prior Work

There are reports in the literature of standard-agnostic approaches similar to ours, which enable a semantically interoperable clinical data landscape. Gaudet-Blavignac et al...
[27] propose a 3-pillar strategy based on a multidimensional encoding of concepts, a resource description framework (RDF)–based storage and transport of the instances of these concepts, and a conversion of the RDF to any target data model. Likewise, the INFOBANCO project of the Madrid Region in Spain [28] aims to create a platform for the management, persistence, exchange, and reuse of health data, contemplating 2 types of outputs: interoperability (HL7 FHIR, EN/ISO 13606, Clinical Data Interchange Standards Consortium [CDISC] [29]) and persistence (OpenEHR, i2b2, Observational Medical Outcomes Partnership Common Data Model [OMOP CDM]). It uses a standard-agnostic design that seeks to apply each health information standard for the purpose it was intended to, offering multiple interoperability and exploitation services suited for specific use cases [12]. However, these projects focus on the creation of interoperable platforms for different purposes, but they do not include a strategy for integrating information coming from mobile apps.

Other groups have reported the use of the SMART on FHIR framework to integrate PGHD from mobile apps into EHRs [30-33]. This framework enables medical apps to be written once and run unmodified across different health care IT systems and has proven to be an effective approach for interoperability. FHIR offers operational mechanisms for data exchange, but unlike EN/ISO 13606, it lacks the capacity to build new concepts based on specific requirements [12], which limits its flexibility to adapt to new scenarios.

Strengths and Limitations

There are strengths to this study that are worth mentioning. First, the 3 main software programs used (LinkEHR, Protégé, and Liferay) are open source, which makes our methodology accessible to low-income areas as well as institutions with limited funding for such projects. Moreover, the aforementioned flexibility and standard-agnostic nature of our methodology define a considerable scalability. The knowledge representation can be adjusted to different contexts with little resources, just by creating new archetypes, modeling the clinical concepts, and mapping them to the corresponding structure of EN/ISO 13606. If a different health information standard is to be used, its metamodel must be represented with ontologies, and both the conversion module and the API need to be adjusted.

With a few exceptions, such as the experience reported by Zenteno et al [34], there is limited evidence in the literature regarding the effective communication and integration of EN/ISO 13606–standardized extracts from a mobile app into an EHR. In addition, to the best of our knowledge, ours is the first experience that does so with data coming from a patient mobile app. Given EN/ISO 13606’s advantages over FHIR in terms of modeling and formalization of clinical knowledge and flexibility in the creation of new concepts, our approach proves to be quite innovative in the communication of EHR extracts with mobile apps.

This study also has some limitations. First, even though there is a log file in the server that registers the extracts that are sent, there is no alarm that notifies us when the process is not working. Therefore, this maintenance and update of the system still depends on manual processes. Furthermore, the ontology-based approach requires trained staff and an initial development that involves the allocation of resources in terms of personnel, funds, and time, which can limit the extensibility of the methodology to other contexts.

Next Steps

Regarding next steps of the project, we are in the process of integrating a dashboard into OntoCR, which will display the AI-based predicted variation in the QoL issues according to the interventions carried out by physicians. This will help physicians with their clinical decision-making when evaluating treatment alternatives for breast cancer patients.

Furthermore, we are working on extending the integration of extracts to other functionalities in Xemio Research, and later, we plan to do so with other mobile apps used within the HCB ecosystem.

Conclusion

This study describes a novel methodology for the successful communication of standardized EHR extracts from a patient mobile app with an ontology-based clinical repository linked to an EHR. Its flexibility and standard-agnostic nature provide significant scalability to adapt to different contexts, situations, and information systems, while the use of open source software facilitates its transferability to other institutions. Our approach allows the integration of data coming from different sources into HISs for them to be used in the caregiving process (primary use) or for investigation purposes (secondary use). To the best of our knowledge, this is the first study to achieve effective communication and integration of EN/ISO 13606–standardized extracts from a patient mobile app into an EHR.

Acknowledgments

This research work was carried out as part of the EU-funded Research and Innovation Action, Artificial intelligence Supporting CAncer Patients across Europe (ASCAPE; Project ID: 875351), [H2020-SC1-DTH-2019] SC1-DTH-01-2019, Big data and Artificial Intelligence for monitoring health status and quality of life after the cancer treatment.

The improvement of the Xemio Research platform was supported by a La Caixa Foundation grant (LCF/PR/AR19/51450002).

Conflicts of Interest

None declared.
References


Abbreviations

- **AI**: artificial intelligence
- **API**: Application Programming Interface
- **ASCAPE**: Artificial intelligence Supporting CAncer Patients across Europe
- **CDISC**: Clinical Data Interchange Standards Consortium
- **EHR**: electronic health record
- **EN/ISO**: European Norm/International Organization for Standardization
- **FHIR**: Fast Healthcare Interoperability Resources
- **FL**: federated learning
- **GUI**: graphical user interface
- **HCB**: Hospital Clínico de Barcelona
- **HIS**: health information system
- **HL7**: Health Level 7
- **ICD-10**: International Classification of Diseases 10th Revision
- **IT**: information technology
- **LOINC**: Logical Observation Identifiers Names and Codes
- **ML**: machine learning
- **OMOP CDM**: Observational Medical Outcomes Partnership Common Data Model
- **PGHD**: patient-generated health data
- **PROM**: patient-reported outcome measure
- **QoL**: quality of life
- **SMART**: Substitutable Medical Applications and Reusable Technologies
- **SNOMED CT**: Systematized Nomenclature of Medicine – Clinical Terms
- **URI**: uniform resource identifier
Successful Integration of EN/ISO 13606–Standardized Extracts From a Patient Mobile App Into an Electronic Health Record: Description of a Methodology


JMIR Med Inform 2022;10(10):e40344
URL: https://medinform.jmir.org/2022/10/e40344
doi: 10.2196/40344
PMID: 36222792

© Santiago Frid, Maria Angeles Fuentes Expósito, Inmaculada Grau-Corral, Clara Amat-Fernandez,Montserrat Muñoz Mateu, Xavier Pastor Duran, Raimundo Lozano-Rubí. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 12.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.
Fast Healthcare Interoperability Resources for Inpatient Deterioration Detection With Time-Series Vital Signs: Design and Implementation Study

Tzu-Wei Tseng¹*, MSc; Chang-Fu Su²*, MD; Feipei Lai¹*, PhD

¹Department of Computer Science and Information Engineering, National Taiwan University, Taipei City, Taiwan
²Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei City, Taiwan
*all authors contributed equally

Abstract

Background: Vital signs have been widely adopted in in-hospital cardiac arrest (IHCA) assessment, which plays an important role in inpatient deterioration detection. As the number of early warning systems and artificial intelligence applications increases, health care information exchange and interoperability are becoming more complex and difficult. Although Health Level 7 Fast Healthcare Interoperability Resources (FHIR) have already developed a vital signs profile, it is not sufficient to support IHCA applications or machine learning–based models.

Objective: In this paper, for IHCA instances with vital signs, we define a new implementation guide that includes data mapping, a system architecture, a workflow, and FHIR applications.

Methods: We interviewed 10 experts regarding health care system integration and defined an implementation guide. We then developed the FHIR Extract Transform Load to map data to FHIR resources. We also integrated an early warning system and machine learning pipeline.

Results: The study data set includes electronic health records of adult inpatients who visited the En-Chu-Kong hospital. Medical staff regularly measured these vital signs at least 2 to 3 times per day during the day, night, and early morning. We used pseudonymization to protect patient privacy. Then, we converted the vital signs to FHIR observations in the JSON format using the FHIR Extract Transform Load application. The measured vital signs include systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, and body temperature. According to clinical requirements, we also extracted the electronic health record information to the FHIR server. Finally, we integrated an early warning system and machine learning pipeline using the FHIR RESTful application programming interface.

Conclusions: We successfully demonstrated a process that standardizes health care information for inpatient deterioration detection using vital signs. Based on the FHIR definition, we also provided an implementation guide that includes data mapping, an integration process, and IHCA assessment using vital signs. We also proposed a clarifying system architecture and possible workflows. Based on FHIR, we integrated the 3 different systems in 1 dashboard system, which can effectively solve the complexity of the system in the medical staff workflow.

(JMIR Med Inform 2022;10(10):e42429) doi:10.2196/42429

KEYWORDS
Fast Healthcare Interoperability Resources; FHIR; Health Level 7; HL7; health research; data sharing; health information technology; clinical research
**Introduction**

**Background**

Vital signs have been an important indicator in many studies [1-3]. In recent years, researchers have used these data in studies of predictive models for in-hospital cardiac arrest (IHCA) [1,4]. In a real-world medical workflow, complete data may be obtained once every 4 to 8 hours. In the machine learning research related to vital signs [5], the features include heart rate, temperature, respiratory rate, systolic blood pressure, and diastolic blood pressure. In addition to IHCA risk assessment, data analysis systems [6] and early warning systems [7] are still indispensable applications.

Although IHCA risk indicators have facilitated breakthroughs in machine learning [8,9], it has been difficult to integrate them into the workflow of medical staff. In hospitals, there are dozens of systems that must exchange information with each other. Without a standard exchange interface [10], the integration process is costly and time-consuming when a new application is imported. In addition, if medical researchers are allowed to access patient data directly through the health care information system database, security risks [11] become a concern.

To begin initiating a human-readable and user-friendly interface for medical staff, Health Level 7 [12] developed Fast Healthcare Interoperability Resources (FHIR) [13]. FHIR is a platform specification that defines a set of capabilities used across the health care processes, and it defines a generic health care business entity model that uses resources as the basic blocks. Each resource in FHIR has a defined relationship resource with data elements and constraints. In addition, the FHIR profile standardizes the data format and structure constraints. During data transportation, it uses the HTTP RESTful application programming interface (API) in the exchange interface and provides the flexibility to choose between JSON or XML format in the data payload.

**Aim**

Although FHIR have developed some of the resources, a vital signs profile [14] has not yet matured. The current implementation guide provided by FHIR is insufficient to encompass the full range of medical system applications; therefore, hospitals still need to define the customized implementation guide to develop their system and workflow. The implementation guide is a collection of rules applied by FHIR resources [15] that requires a clear explanation of how to solve a particular problem. In the relevant studies on FHIR [16-18], each paper develops and discusses a single customized resource profile on a mobile device. Although FHIR can effectively and rapidly improve health care information system interoperability, it still has not proposed an implementation guide for the machine learning application in FHIR implementation guide registry. To accelerate the development of smart health care, we propose a system architecture process based on FHIR that can integrate the machine learning models. Besides, the vital signs applications are distributed in many different systems. This study can effectively solve the complexity of the system in the medical staff workflow.

To standardize the format among medical systems, we developed a complete IHCA implementation guide based on FHIR that defines the vital signs–related data for both the early warning system and the machine learning pipeline. In addition, we also developed FHIR Extract Transform Load (ETL) and other FHIR-related applications, including data management, an early warning system, and a machine learning pipeline.

**Methods**

**Ethics Approval**

This study was approved by the Institutional Review Board of the En-Chu-Kong Hospital (ECKIRB1071001). We confirm that all experiments were performed in accordance with relevant guidelines and regulations. The data retrieved from electronic health records (EHRs) were deidentified by an IT specialist and could not be linked to the patients’ identity by the research team. The need for written informed consent was waived and confirmed by the En-Chu-Kong Hospital Institutional Review Board, because this was a retrospective cohort study with deidentified data.

**Overview**

Our study provides a design and implementation process for IHCA-based interoperability of health care information systems, and our design steps include use cases as well as the IHCA implementation guide.

In the use cases section, we describe the integration issues faced by health care institutions. Then, in the IHCA implementation guide section, we introduce the method used to migrate data from the healthcare information system (HIS) database to the FHIR server as well as a method for mapping the data to the FHIR resources. We also develop the 3 application systems, which include data management, early warning systems, and a machine learning pipeline. If used according to our implementation guide, the applications can easily obtain patient information and vital signs data.

**Use Case Survey**

In health care institutions, the database is centrally managed, but the applications are developed by many different teams. In addition, medical staff usually access all of the required information about a workflow through a single system. Therefore, the interoperability of health care systems is very important.

To achieve system information interoperability [19], the HTTP RESTful API was defined to exchange data with other systems. However, many medical systems are legacy systems, and in many cases, it is impossible to change the system architecture. We therefore created a table view for the HIS database to allow other systems to obtain particular data. To avoid affecting the original system architecture, we developed FHIR ETL to convert data from the HIS database to the FHIR server, and FHIR ETL was implemented according to the rules defined by the IHCA implementation guide.

We interviewed 10 experts regarding health care system integration and information exchange. As shown in Table 1, FHIR, which has a good medical standard interface, is very
suitable for solving the interoperability problems faced by medical information systems. In addition, it supports a variety of systems that can be used to develop extended applications.

Therefore, we have 2 use cases. The first use case is related to data migration for the FHIR server, as shown in Figure 1 (Part A). The second use case is related to FHIR applications, as shown in Figure 1 (Part B).

### Table 1. Requirement list from health care specialists in health care institutions.

<table>
<thead>
<tr>
<th>Issue (requirement)</th>
<th>How to do it</th>
</tr>
</thead>
<tbody>
<tr>
<td>The new system integration process shall not affect the health care information system or the vital signs system.</td>
<td>Build the FHIR server as a new middleware or gateway so that researchers can access data.</td>
</tr>
<tr>
<td>Converting the EHRs with vital signs into FHIR resources.</td>
<td>Develop the FHIR ETL.</td>
</tr>
<tr>
<td>To reduce the time cost and compatibility, we need to use a health care information interoperability standard.</td>
<td>Use FHIR resources and the RESTful API.</td>
</tr>
<tr>
<td>The field needs an early warning system that can continuously monitor the patient’s vital signs.</td>
<td>Use FHIR to develop the early warning system.</td>
</tr>
<tr>
<td>How can an organization integrate the prediction model into the medical workflow?</td>
<td>Use FHIR to develop the machine learning pipeline.</td>
</tr>
<tr>
<td>The field needs a complete implementation procedure and use case.</td>
<td>Define an FHIR implementation guide.</td>
</tr>
</tbody>
</table>

*a* FHIR: Fast Healthcare Interoperability Resources.  
*b* EHRs: electronic health records.  
*c* ETL: Extract Transform Load.  
*d* API: application programming interface.

Figure 1. Use cases for IHCA research and application. (A) Extract the data and transfer them to the FHIR server. (B) Data management for data processing, early warning system for notification and model trigger, and machine learning pipeline for model prediction and model training. API: application programming interface; ETL: Extract Transform Load; FHIR: Fast Healthcare Interoperability Resources; HAPI: Health Level 7 application programming interface; HIS: healthcare information system; IHCA: in-hospital cardiac arrest.

### IHCA Implementation Guide

In this phase, we need to consider the data format so that raw data can be transferred into FHIR resources as well as how the HTTP RESTful API can be used to easily obtain data. Therefore, we designed a system architecture (Figure 1). We divided the system steps into the following: (1) the FHIR ETL performs data conversion and comparisons between the HIS database and the FHIR server, and (2) the application system accesses data directly through the FHIR API interface at the HTTP layer.

### Data Mapping—FHIR ETL

We proposed the data mapping table to develop the FHIR ETL, as shown in Table 2. We defined the data mapping and resource relations. Based on the FHIR vital signs profile, we used the observation resource to store systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, and body temperature. According to Table 2, FHIR ETL can extract the data from the HIS database and transfer them to resource content.
### Table 2. The data mapping table of the FHIR\(^a\) ETL\(^b\) in this study.

<table>
<thead>
<tr>
<th>Data model of HIS(^c) database</th>
<th>FHIR resource name</th>
<th>FHIR resource attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient_ID</td>
<td>Patient</td>
<td>identifier</td>
<td>An identifier for the patient in the hospital</td>
</tr>
<tr>
<td>Patient_name</td>
<td>Patient</td>
<td>name</td>
<td>Patient’s name that is human-readable</td>
</tr>
<tr>
<td>Gender</td>
<td>Patient</td>
<td>gender</td>
<td>Patient’s gender</td>
</tr>
<tr>
<td>BirthDate</td>
<td>Patient</td>
<td>birthDate</td>
<td>Patient’s birth date</td>
</tr>
<tr>
<td>Practitioner_ID</td>
<td>Practitioner</td>
<td>identifier</td>
<td>An identifier for the physician in the hospital</td>
</tr>
<tr>
<td>Practitioner_name</td>
<td>Practitioner</td>
<td>name</td>
<td>Physician’s name that is human-readable</td>
</tr>
<tr>
<td>Organization_ID</td>
<td>Organization</td>
<td>identifier</td>
<td>An identifier for the department in the hospital</td>
</tr>
<tr>
<td>Organization_name</td>
<td>Organization</td>
<td>name</td>
<td>Department’s name that is human-readable</td>
</tr>
<tr>
<td>Location_ID</td>
<td>Location</td>
<td>identifier</td>
<td>An identifier for the location in the hospital</td>
</tr>
<tr>
<td>Location_name</td>
<td>Location</td>
<td>name</td>
<td>Location’s name that is human-readable</td>
</tr>
<tr>
<td>Heart rate</td>
<td>Observation</td>
<td>valueQuantity.value</td>
<td>Heart rate</td>
</tr>
<tr>
<td>Temperature</td>
<td>Observation</td>
<td>valueQuantity.value</td>
<td>Temperature</td>
</tr>
<tr>
<td>Respiratory rate</td>
<td>Observation</td>
<td>valueQuantity.value</td>
<td>Respiratory rate</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>Observation</td>
<td>valueQuantity.value</td>
<td>Systolic blood pressure</td>
</tr>
<tr>
<td>Diastolic blood pressure</td>
<td>Observation</td>
<td>valueQuantity.value</td>
<td>Diastolic blood pressure</td>
</tr>
<tr>
<td>Timestamp</td>
<td>Observation</td>
<td>effectiveDateTime</td>
<td>The created time of the value</td>
</tr>
</tbody>
</table>

\(^a\)FHIR: Fast Healthcare Interoperability Resources.  
\(^b\)ETL: Extract Transform Load.  
\(^c\)HIS: healthcare information system.

In Figure 2, in terms of data acquisition, if an FHIR client wants to obtain a patient’s location, it needs to first obtain the patient’s ID and join the encounter subject. Then, it can use the encounter location to find the location resource. Finally, the FHIR client can obtain the patient location.

In Figure 3, the FHIR client can perform the following: (1) when an FHIR client needs to access a particular patient using metadata, it can use the HTTP GET method to obtain the Bundle resource response; (2) when an FHIR client wants to update the location name for the hospital, it can use the HTTP PUT method to update the Location resource; and (3) after the FHIR client obtains sufficient vital signs data from the Observation resource, it can use the HTTP DELETE method to delete the resource that is missing vital signs values.
Figure 2. Data mapping and resource relationships in the IHCA implementation guide. ETL: Extract Transform Load; FHIR: Fast Healthcare Interoperability Resources; HIS: healthcare information system; IHCA: in-hospital cardiac arrest.

Figure 3. FHIR (Fast Healthcare Interoperability Resources) application, which uses the HTTP RESTful API (application programming interface) to control the data on the FHIR server.

Workflow Design

In this section, we describe the complete workflow of FHIR implementation. Workflow 1 develops the data mappings for the FHIR resources. First, the FHIR ETL uses the database connection library to access the table view of the HIS database. Then, it verifies that the patient’s information exists. To maintain data consistency, when converting to the Observation resource, the system must add the universally unique identifier of Patient resource as a reference link. If the patient’s basic data already exists, the vital signs will be converted into an Observation resource based on the FHIR vital signs profile.

Workflow 2 develops the data acquisition process for FHIR applications. First, the FHIR application can use URL (/Patient)
with the HTTP GET method to access the Bundle resource. In
the Bundle resource, the FHIR application can find all of the
patient’s data. If the FHIR application needs to obtain patient
information such as location and practitioner information, it can
use the Patient ID to join the Encounter subject. Then, it can
obtain the Encounter participant and Encounter location. Finally,
it can also use the Patient ID to join the URL
(\Observation?subject=) with the HTTP GET method to obtain
the Observation resource (Figure 4).

Figure 4. Workflow of the Fast Healthcare Interoperability Resources (FHIR) Extract Transform Load (ETL) and the FHIR client application. HIS:
healthcare information system; UUID: universally unique identifier.

Results

FHIR Resources

The FHIR ETL is an automation service that extracts vital signs.
When the vital signs system stores the data in the HIS database,
the FHIR ETL can access the vital signs data immediately, and
as shown in Figure 2, it adds the vital signs to the Observation
resource. Multimedia Appendix 1 shows examples of an FHIR
resource that refers to an FHIR vital signs profile and other
resources.

Software Development

We describe the software development, which is shown in
Figure 5. The HIS database was developed using the SQL server
database and the Oracle database server. The FHIR server was
installed on the Health Level 7 API FHIR R4 server (version
6.1.0) [20] with a docker container based on the Java
environment. This open-source system is widely used. We
developed the back-end service of the FHIR ETL using Python
software (version 3; Python Software Foundation), and the
machine learning pipeline was implemented using Flask. The
front-end website was constructed using Vue.js and is used for
IHCA web management.
Figure 5. System architecture used in this study. API: application programming interface; ETL: Extract Transform Load; FHIR: Fast Healthcare Interoperability Resources; HAPI: Health Level 7 application programming interface; HIS: healthcare information system; IHCA: in-hospital cardiac arrest; MLP: machine learning pipeline.

System Implementation

The study data set [21] included the EHRs of adult inpatients who visited the En-Chu-Kong hospital. Medical staff regularly measured these vital signs at least 2 to 3 times per day during the day, night, and early morning. The total number of patients was 16,865, and the number of patients with IHCA was 118.

We converted the 5 vital signs into FHIR observations in JSON format using FHIR ETL. These vital signs include systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, and body temperature. For demonstration, we used pseudonymization [22] to protect the patient’s privacy. Furthermore, we divided the proposed system into the following 3 components: data management, an early warning system, and a machine learning pipeline. In terms of data management, as shown in Figure 6, we developed a data static dashboard so that it can be accessed by medical staff using a browser. The dashboard uses the HTTP GET method to obtain both the Patient and Observation resources. Then, the patient’s vital signs over the previous 48 hours are displayed. In the early warning system, medical staff can set the vital signs alert threshold to decide whether to show the alert in the notification list as shown in Figure 7. Then, the machine learning pipeline exports the vital signs data from the Observation resource to the FHIR server. We integrated a long short-term memory network–based model [21] using vital signs data to predict IHCA. It used the time series early warning score, which used heart rate, systolic blood pressure, and respiratory data. When the training process of the prediction model is initiated, the status “in progress” will appear in MongoDB. After model training, the status will be updated to “final,” and the dashboard will show the latest accuracy of the model. The proposed dashboard is shown in Figure 8. However, the system can be used on mobile devices as well as desktop computers. We followed the Responsive Web Design [23] to design a user-friendly mobile interface (Figure 9).
Figure 6. Screenshot of the data management overview in the dashboard.

Figure 7. Screenshot of the early warning system’s notification overview.
Discussion

Principal Findings

In this paper, we piloted the use of an implementation guide that combines IHCA with vital signs, which have been widely adopted in IHCA assessment [4,21] and play an important role in inpatient deterioration detection. Many health care institutions have developed early warning score systems to identify hospitalized patients that are at risk of deterioration, and in recent years, they have begun to incorporate machine learning–based models into this process. To promote system interoperability, we used the FHIR standard to achieve consistent information exchange. We also combined 5 resources (Organization, Location, Practitioner, Patient, and Encounter) to represent the EHR. Then, based on the FHIR vital signs profile, we exported vital signs data to HIS database and defined the IHCA implementation. In addition, we developed the 3 FHIR applications of data management using a dashboard, a real-time early warning system, and a machine learning–based pipeline. According to the IHCA implementation guide, our proposed system makes it easy to integrate vital signs–related applications.

Figure 8. Screenshot of the machine learning pipeline including prediction and training. DEP: department; IHCA: in-hospital cardiac arrest.

Figure 9. User interface developed using Responsive Web Design for mobile devices. IHCA: in-hospital cardiac arrest.
Limitations

The implementation guide was only developed for vital signs-related studies. However, some case studies still need to include treatment history [24], blood urea nitrogen [25], and creatinine [25]. These further improvements can be made to the EHR.

Comparison With Prior Work

Despite the result that indicated that FHIR can improve the interoperability of health care information systems [26-28], existing studies have only developed the resource and profiles. Seong et al [16] demonstrated how quality information regarding clinical next-generation sequencing genomic testing can be exchanged in a standardized format by profiling an FHIR genomic resource and developing an FHIR-based web application that exchanges quality information. Based on the human-centered design methodology, Park et al [17] developed a worker-centered personal health record (PHR) app for occupational health. The PHRs were managed through a cloud server using Azure API for FHIR, and the PHR FHIR resources included Patient, Organization, DiagnosticReport, Observation, Practitioner, Condition, Procedure, MedicationStatement, Medication, and Encounter. In addition, Chukwu et al [18] profiled FHIR resources for maternal and child health referrals.

Conclusions

We successfully demonstrated a process that standardizes health care information for inpatient deterioration detection using vital signs. Based on the FHIR definition, we provided an implementation guide that includes data mapping, an integration process, and IHCA assessment using vital signs. We also provided a clarified system architecture that can be used to develop clinical decision support systems. Based on FHIR, we integrated the 3 different systems into 1 dashboard system, which can effectively solve the complexity of the system in the medical staff workflow.

Acknowledgments

This paper was partly supported by the Ministry of Science and Technology, Taiwan (grant 10X-62634-F-002-015). The authors acknowledge the support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

All Health Level 7 Fast Healthcare Interoperability Resources Bundle responses in this study.

References


Abbreviations

API: application programming interface
EHR: electronic health record
ETL: Extract Transform Load
FHIR: Fast Healthcare Interoperability Resources
HIS: healthcare information system
IHCA: in-hospital cardiac arrest
PHR: personal health record

Edited by M Focs; submitted 04.09.22; peer-reviewed by A Nassirian, R Saripalle, T Zhang; comments to author 20.09.22; revised version received 22.09.22; accepted 03.10.22; published 13.10.22.

Please cite as:
Tseng TW, Su CF, Lai F
Fast Healthcare Interoperability Resources for Inpatient Deterioration Detection With Time-Series Vital Signs: Design and Implementation Study
JMIR Med Inform 2022;10(10):e42429
URL: https://medinform.jmir.org/2022/10/e42429
doi:10.2196/42429
PMID:36227686

©Tzu-Wei Tseng, Chang-Fu Su, Feipei Lai. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 13.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.
Coronary Artery Computed Tomography Angiography for Preventing Cardio-Cerebrovascular Disease: Observational Cohort Study Using the Observational Health Data Sciences and Informatics’ Common Data Model

Woo Kyung Bae¹*, MPH, MD; Jihoon Cho²*, MS; Seok Kim², MPH; Borham Kim², BSN, RN; Hyunyoung Baek², MPH, RN; Wongeun Song², MS; Sooyoung Yoo², PhD
1Department of Family Medicine, Health Promotion Center, Seoul National University Bundang Hospital, Republic of Korea, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea
2Healthcare Information and Communication Technology Research Center, Office of eHealth Research and Business, Seoul National University Bundang Hospital, Republic of Korea, Seongnam-si, Republic of Korea
*these authors contributed equally

Corresponding Author:
Sooyoung Yoo, PhD
Healthcare Information and Communication Technology Research Center, Office of eHealth Research and Business
Seoul National University Bundang Hospital
Republic of Korea
172, Dolma-ro, Bundang-gu
Seongnam-si, 13605
Republic of Korea
Phone: 82 010 9053 7094
Email: yoosoo0@snubh.org

Abstract

Background: Cardio-cerebrovascular diseases (CVDs) result in 17.5 million deaths annually worldwide, accounting for 46.2% of noncommunicable causes of death, and are the leading cause of death, followed by cancer, respiratory disease, and diabetes mellitus. Coronary artery computed tomography angiography (CCTA), which detects calcification in the coronary arteries, can be used to detect asymptomatic but serious vascular disease. It allows for noninvasive and quick testing despite involving radiation exposure.

Objective: The objective of our study was to investigate the effectiveness of CCTA screening on CVD outcomes by using the Observational Health Data Sciences and Informatics’ Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) data and the population-level estimation method.

Methods: Using electronic health record–based OMOP-CDM data, including health questionnaire responses, adults (aged 30-74 years) without a history of CVD were selected, and 5-year CVD outcomes were compared between patients undergoing CCTA (target group) and a comparison group via 1:1 propensity score matching. Participants were stratified into low-risk and high-risk groups based on the American College of Cardiology/American Heart Association atherosclerotic cardiovascular disease (ASCVD) risk score and Framingham risk score (FRS) for subgroup analyses.

Results: The 2-year and 5-year risk scores were compared as secondary outcomes between the two groups. In total, 8787 participants were included in both the target group and comparison group. No significant differences (calibration \( P=.37 \)) were found between the hazard ratios of the groups at 5 years. The subgroup analysis also revealed no significant differences between the ASCVD risk scores and FRSs of the groups at 5 years (ASCVD risk score: \( P=.97 \); FRS: \( P=.85 \)). However, the CCTA group showed a significantly lower increase in risk scores at 2 years (ASCVD risk score: \( P=.03 \); FRS: \( P=.02 \)).

Conclusions: Although we could not confirm a significant difference in the preventive effects of CCTA screening for CVDs over a long period of 5 years, it may have a beneficial effect on risk score management over 2 years.

(JMIR Med Inform 2022;10(10):e41503) doi:10.2196/41503
KEYWORDS
cardiovascular diseases; coronary artery computed tomography angiography; observational study; common data model; population level estimation; cardiology; vascular disease; medical informatics; computed tomography; angiography; electronic health record; risk score; health data science; data modeling

Introduction
Cardio-cerebrovascular diseases (CVDs) result in 17.5 million deaths annually worldwide, accounting for 46.2% of noncommunicable causes of death, and are the leading cause of death, followed by cancer, respiratory disease, and diabetes mellitus [1]. CVDs involve demographic factors (age, sex, and family history), pre-existing conditions (hypertension, diabetes mellitus, and hyperlipidemia), and lifestyle and environmental factors. Unlike demographic characteristics, lifestyle factors, such as an inappropriate diet, a lack of exercise, smoking, stress, and excessive drinking, can be improved to reduce the risk of CVDs [2].

Coronary artery computed tomography angiography (CCTA) detects calcification in the coronary arteries and can be used to detect asymptomatic but serious vascular disease. It allows for noninvasive and quick testing despite involving radiation exposure [3,4]. For these reasons, many studies have investigated the early detection of CVDs by using CCTA, which enables prompt treatment and results in better outcomes.

In recent years, there has been debate about whether screening via CCTA helps prevent CVDs in populations with varying degrees of risk. CCTA has been recommended to predict CVDs in patients with cancer [2,5], but among asymptomatic individuals, the evidence about its effectiveness is inconsistent.

We aimed to study the effectiveness of CCTA screening by analyzing observational health checkup data from electronic health records (EHRs) in the form of the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), using a cohort study design [6]. The OMOP-CDM standardizes disparate data and enables the analysis of deidentified, large-scale observational data in a distributed research data network. Moreover, as the data are standardized, the same analytical codes can be used to conduct efficient analyses through the data network. Observational Health Data Sciences and Informatics (OHDSI)—an open international collaborative community—provides an open-source analytics tool for OMOP-CDM data that produces scientific, reliable, and reproducible evidence.

Using the OHDSI analytics tool, we performed a comparative effectiveness study of CVD outcomes in asymptomatic patients without a history of CVD who underwent a health checkup at a tertiary university hospital. The conventional assessments of CVD risk, namely assessments of the Framingham risk score (FRS) and the American College of Cardiology/American Heart Association (ACC/AHA) atherosclerotic cardiovascular disease (ASCVD) risk score, were used to stratify the participants into high-risk and low-risk groups for stratified analyses. Although the risk of CVD increases with age, we compared differences between the two groups after 2 and 5 years to assess the short-term benefits of CCTA-based screening and whether it can help prevent CVDs.

Methods

Data Sources
The study site was the Seoul National University Bundang Hospital (SNUBH), which is located in the Seoul metropolitan area. The SNUBH collected OMOP-CDM version 5.3 data based on EHRs from 2003 to 2020. The data included patients’ demographic information, clinical information (diagnoses, medications, tests, surgeries and procedures, family histories, past histories, and nursing flowcharts), and health questionnaire responses. The health questionnaire responses about medical history, family history, socioeconomic status, medication history, marital status, exercise and physical activity status, and depression assessment results were converted to OMOP-CDM data. In this study, we used the deidentified OMOP-CDM data that the SNUBH collected from over 2 million patients, including outpatients, inpatients, and emergency department visits.

Ethical Considerations
This study adhered to the relevant guidelines and regulations of the SNUBH Institutional Review Board (IRB). As the OMOP-CDM is a deidentified data set, the study was exempted from review by the SNUBH IRB (IRB number: X-2202-736-903).

Study Design
This was a retrospective, observational, comparative cohort study that used OMOP-CDM–formatted EHR data. We analyzed data from adults aged 30 to 74 years who underwent a health checkup between April 1, 2003, and December 31, 2015, and were followed up for at least 5 years. Only those who responded to the questionnaire item about medical history in the health checkup survey were included. Individuals with a history of CVD were excluded from this study. The index date was set as the date of completing the health checkup questionnaire at a health checkup visit for the first time. CVDs that occurred within 60 days of the index date were considered as cases in which patients were diagnosed during the health checkup, and these CVD events were excluded as CVD outcomes. Thus, the outcome was defined as CVD events that occurred 60 days after the index date, and follow-ups ended on the date that CVD events occurred (ie, within 5 years from the index date), the date of the final hospital visit, or the date of death. As such, the time-at-risk period was set as 61 days after the index date to 5 years after the index date.

The primary outcome was the comparison of CVD hazard ratios (HRs) between the group that underwent CCTA (target group) and the group that did not undergo CCTA at the health checkup visit (comparison group).

In the subgroup analyses, the CVD HRs, which were based on the ACC/AHA ASCVD risk score and the FRS, were analyzed. The patients were stratified into the nonrisk and low-risk group
or the high-risk group based on a cutoff score of 10 for the FRS [7] and 5 for the ASCVD risk score [8].

The secondary outcome was the difference between the risk scores of patients who underwent health checkups 2 years and 5 years after the index date. The differences between the risk scores at the index date and those at the times of subsequent examinations were used for comparative analyses.

**Study Population**

From April 2003 to December 2015, a total of 69,334 patients aged 30 to 74 years were enrolled for a health checkup. Of these patients, only 49,496 responded to the questionnaire, and only 46,087 patients had no cardiovascular history. A total of 42,489 patients for whom we could calculate the risk score—a key indicator of this study—were selected as the initial cohort.

Initially, of the 42,489 patients who were included in the analysis, 12,661 underwent CCTA (target group), and 29,828 did not (comparison group). Of these patients, 1514 from the target group and 1519 from the comparison group with a history of CVD before the index date were excluded from the analysis. In addition, 1783 patients from the target group and 5004 patients from the comparison group who did not fulfill the minimum observation period of 1 day during the time-at-risk window were excluded. The remaining 9364 patients from the target group and 23,305 patients from the comparator group underwent 1:1 propensity score matching. During 1:1 propensity score matching, 577 people who did not match the comparator group were excluded from the target group because matching was performed to maximize the minority group, and 14,518 people were excluded from the comparator group. Finally, 8787 of the 12,661 patients (69.4%) from the initial target cohort were selected as the final target group, and 8787 of the 29,828 patients (29.5%) from the initial comparator cohort were used for the analysis as the final comparator group (Figure 1).
**Figure 1.** The flowchart of the study population. CCTA: coronary artery computed tomography angiography.

**Covariates**

Approximately 13,000 variables were used as covariates for propensity score matching. These covariates included patient clinical data that were obtained at any time prior to the index date and health checkup data that were obtained on the index date. The patient clinical covariates included the condition era, the condition group era, the drug group era, observations, measurements, procedures, the Charlson Comorbidity Index score, the Diabetes Complications Severity Index score, the CHADS<sub>2</sub> (Congestive Heart Failure, Hypertension, Age, Diabetes, Previous Stroke/Transient Ischemic Attack [2 points]) score, the CHA<sub>2</sub>D<sub>S</sub>VASc (Congestive Heart Failure, Hypertension, Age≥75 [Doubled], Diabetes, Stroke [Doubled], Vascular Disease, Age 65 to 74, and Sex Category [Female]) score, and the hospital frailty risk score. The covariates that were measured at the index date included demographic data, such as sex, age, education level, average monthly income, and
marital status; health questionnaire data, such as any history of cancer and chronic diseases (hypertension, diabetes, and hyperlipidemia), medication history (antihypertensive drugs, antidiabetic drugs, antihyperlipidemic drugs, and aspirin), smoking status, and family history; and health checkup data, such as height, weight, BMI, blood pressure (systolic and diastolic), waist circumference, glucose levels, uric acid levels, aspartate aminotransferase levels, alanine aminotransferase levels, triglyceride levels, total cholesterol, high-density lipoprotein cholesterol levels, low-density lipoprotein cholesterol levels, and glycated hemoglobin A1c levels.

Outcomes
The outcome of this study was the first registered CVD event, which was based on a CVD diagnosis during the observation period. A CVD event was defined based on International Classification of Diseases, 10th Revision (ICD-10) codes I20 to I25 (ischemic heart disease), I50 (heart failure), I60 to I69 and G45 to G46 (stroke), and E78 (hypercholesterolemia). As we intended to assess the HRs of CVDs resulting from arteriosclerotic diseases only, we excluded cardiogenic diseases, such as atrial fibrillation and aneurysm (I42-I43, I48, I71, I62, and I68), and diseases caused by external accidental factors (I60 and I62). The ICD-10 codes that were chosen as the outcomes were reviewed by 1 clinical specialist and 1 nurse.

Statistical Analysis
We used the population-level estimation methodology and an open-source tool provided by OHDSI [9]. All analyses were performed by using R version 4.0.3 (R Foundation for Statistical Computing) [10]. Large-scale propensity score matching [11] was performed to adjust for potential confounding and to resolve the imbalance between the target and comparison cohorts caused by selection bias—a result of the retrospective observational nature of this study. The propensity score–matched model, which used approximately 13,000 covariates, was fitted through regularized regression, and the propensity score was calculated as the probability of a patient undergoing CCTA based on the covariates. Target and comparison group patients with similar propensity scores were matched to create a balanced cohort. To establish a matched cohort, we performed 1:1 propensity score matching by using a caliper width of 0.2 of the SD of the logit. The conditional Cox proportional hazards model was used to estimate HRs for the target group, in relation to the comparison group. The balance of the covariates between the cohorts was assessed based on the standardized difference of the mean (<0.1). Statistical significance was evaluated at $P<.05$ for 2-tailed tests.

To explain any residual bias after controlling for the measured covariates, we used negative control outcomes that were unlikely to be induced or prevented by undergoing CCTA; thus, the actual HR was anticipated to be 1. The negative control outcomes were selected by a clinical specialist through a manual review of the outcomes that were used in a previous OHDSI study [12] (Table S1 in Multimedia Appendix 1). The same study design was used to estimate the outcomes of interest and calculate the HR estimate for the negative control group, and all HR estimates were presented with 95% CIs and $P$ values, along with the empirical null distribution and adjustment [13,14]. The empirical equivalence of the two cohorts was assessed by using the propensity score distribution. We also reported the power analysis; propensity score; cohort balance before and after propensity score matching; fitted null distribution; calibration chart for negative control outcomes; and Kaplan-Meier curve, which shows the proportional hazards assumption over time.

To confirm the changes in the differences in ASCVD risk scores and FRSs, we used the 2-group comparison method. The normality of the amount of change was confirmed by using the Shapiro-Wilk test, and the changes in the two groups were confirmed by using the Wilcoxon rank-sum test.

Results

Characteristics of Study Participants
Table 1 shows the baseline characteristics of the patients before and after propensity score matching. The table shows the patients’ age groups, sex, and BMIs; the number of patients in the risk score groups; and the follow-up periods. For most demographic characteristics, the differences between groups decreased after matching. The standardized difference of the mean for the covariates decreased from 0.4 to 0.07 after propensity score matching, which is lower than the conventional standard of 0.1, thereby confirming that propensity score matching was performed correctly (Figure 2). This can also be observed in Figure 3, which compares the distributions from before and after propensity score matching.

https://medinform.jmir.org/2022/10/e41503
Table 1. The baseline characteristics of the study population before and after propensity score matching.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Before matching</th>
<th></th>
<th>After matching</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCTA group (n=12,661)</td>
<td>Non-CCTA group (n=29,828)</td>
<td>Standard difference</td>
<td>CCTA group (n=8787)</td>
</tr>
<tr>
<td>Age group (years), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-34</td>
<td>226 (1.8)</td>
<td>2442 (8.2)</td>
<td>−0.26</td>
<td>150 (1.7)</td>
</tr>
<tr>
<td>35-39</td>
<td>1043 (8.2)</td>
<td>4319 (14.5)</td>
<td>−0.19</td>
<td>761 (8.7)</td>
</tr>
<tr>
<td>40-44</td>
<td>1870 (14.8)</td>
<td>5257 (17.6)</td>
<td>−0.08</td>
<td>1406 (16)</td>
</tr>
<tr>
<td>45-49</td>
<td>2516 (19.9)</td>
<td>5134 (17.2)</td>
<td>0.07</td>
<td>1846 (21)</td>
</tr>
<tr>
<td>50-54</td>
<td>2435 (19.2)</td>
<td>4617 (15.5)</td>
<td>0.10</td>
<td>1702 (19.4)</td>
</tr>
<tr>
<td>55-59</td>
<td>2084 (16.5)</td>
<td>3322 (11.1)</td>
<td>0.16</td>
<td>1373 (15.6)</td>
</tr>
<tr>
<td>60-64</td>
<td>1468 (11.6)</td>
<td>2275 (7.6)</td>
<td>0.14</td>
<td>908 (10.3)</td>
</tr>
<tr>
<td>65-69</td>
<td>734 (5.8)</td>
<td>1564 (5.2)</td>
<td>0.02</td>
<td>471 (5.4)</td>
</tr>
<tr>
<td>70-74</td>
<td>285 (2.3)</td>
<td>898 (3.0)</td>
<td>−0.05</td>
<td>170 (1.9)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4757 (37.6)</td>
<td>12,650 (42.4)</td>
<td>−0.10</td>
<td>3561 (40.5)</td>
</tr>
<tr>
<td>Male</td>
<td>7904 (62.4)</td>
<td>17,178 (57.6)</td>
<td>0.10</td>
<td>5226 (59.5)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>24.2 (3.1)</td>
<td>23.7 (0.2)</td>
<td>0.18</td>
<td>24.0 (3.1)</td>
</tr>
<tr>
<td>ACC/AHA ASCVD risk score, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (≥5)</td>
<td>5036 (39.8)</td>
<td>8576 (28.8)</td>
<td>N/A</td>
<td>3062 (34.8)</td>
</tr>
<tr>
<td>Low (&lt;5)</td>
<td>7625 (60.2)</td>
<td>21,252 (71.2)</td>
<td>N/A</td>
<td>5725 (65.2)</td>
</tr>
<tr>
<td>Framingham risk score, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (≥10)</td>
<td>4996 (39.5)</td>
<td>8155 (27.3)</td>
<td>N/A</td>
<td>3030 (34.5)</td>
</tr>
<tr>
<td>Low (&lt;10)</td>
<td>7665 (60.5)</td>
<td>21,673 (72.7)</td>
<td>N/A</td>
<td>5757 (65.5)</td>
</tr>
<tr>
<td>Follow-up period (days), mean (SD)</td>
<td>2220.3 (1731.6)</td>
<td>1928.9 (1675.5)</td>
<td>N/A</td>
<td>2604 (1594.4)</td>
</tr>
</tbody>
</table>

aCCTA: coronary artery computed tomography angiography.
bVariables used in propensity score matching.
cACC/AHA: American College of Cardiology/American Heart Association.
dASCVD: atherosclerotic cardiovascular disease.
eVariables not used in propensity score matching.
fN/A: not applicable.
The Cox proportional hazards model was used to estimate and compare the HRs of CVDs among the target and comparison groups after propensity score matching, and no statistically significant differences were found between the two groups. The Kaplan-Meier analysis revealed that the HR was 1.048 (95% CI 0.960-1.144), which was not statistically significant ($P = .30$). The calibration $P$ value, which was adjusted by using a negative control and was the most important indicator in our analysis, was .37, indicating no statistical significance (Figure 4).
Subgroup Analysis

The study population was stratified based on the cutoff scores for the ACC/AHA ASCVD risk score and FRS for subgroup analyses. Table 2 presents the results of each analysis. In each subgroup, the standardized difference of the mean dropped to <0.1 after propensity score matching. Figure S1 in Multimedia Appendix 1 shows the propensity score distributions, and Figure S2 in Multimedia Appendix 1 shows the standardized difference of the mean among groups of covariates before and after propensity score matching.

In the ASCVD high-risk subgroup (risk score ≥ 5), 3149 patients were included in both the target group and comparison group. In the low-risk subgroup (risk score < 5), 5524 patients were included in both the target group and comparison group. In the high-risk and low-risk subgroups, the calibration $P$ value, which was adjusted by using negative controls, was .39 and .50, respectively, showing no significant differences in the HRs of CVDs among the target and comparison groups.

In the FRS high-risk subgroup (FRS ≥ 10), 3110 participants were included in both the target group and comparison group. In the low-risk subgroup (FRS < 10), 5602 patients were included in both the target group and comparison group. The calibration $P$ value, which was adjusted by using negative controls, was .13 and .57 in the high-risk and low-risk subgroups, respectively, indicating no significant differences in the HRs of CVDs among the target and comparison groups (Figure S3 in Multimedia Appendix 1).

Table 2. The risk of cardio-cerebrovascular disease at 5 years in each subgroup based on the American College of Cardiology/American Heart Association (ACC/AHA) atherosclerotic cardiovascular disease (ASCVD) risk score and Framingham risk score (FRS).

<table>
<thead>
<tr>
<th>Risk Score</th>
<th>Hazard ratio (95% CI)</th>
<th>$P$ value$^a$</th>
<th>Calibration $P$ value$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC/AHA ASCVD risk score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (≥ 5)</td>
<td>1.113 (0.984-1.259)</td>
<td>.09</td>
<td>.39</td>
</tr>
<tr>
<td>Low (&lt; 5)</td>
<td>0.999 (0.881-1.133)</td>
<td>.99</td>
<td>.50</td>
</tr>
<tr>
<td>FRS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (≥10)</td>
<td>1.166 (1.031-1.321)</td>
<td>.02</td>
<td>.13</td>
</tr>
<tr>
<td>Low (&lt;10)</td>
<td>1.004 (0.883-1.141)</td>
<td>.96</td>
<td>.57</td>
</tr>
</tbody>
</table>

$^a$Kaplan-Meier analysis $P$ value.
$^b$Calibration $P$ value that was adjusted by using a negative control.

Risk Scores at 2 and 5 Years

The 2-year median change in the ASCVD risk scores and the FRSs of the non-CCTA group was 0.23 and 0.60, respectively. In contrast, the ASCVD risk scores and the FRSs of the CCTA group changed by 0.17 and 0.39, respectively. There was a statistically significant difference for both risk scores, with $P$ values of .03 and .02, respectively.

The 5-year median change in the ASCVD risk scores and the FRSs of the non-CCTA group was 1.06 and 1.61, respectively. In contrast, the ASCVD risk scores and the FRSs of the CCTA group changed by 1.10 and 1.66, respectively. There was no statistically significant difference for both risk scores, with $P$ values of .97 and .85, respectively (Table 3).
Our study compared patients who did or did not undergo additional coronary computed tomography. Both groups underwent the same levels of examination and counseling, which were conducted by the cardiovascular health screening program of the national service in 1 hospital.

Smoking status, blood pressure, and blood lipid concentration, which are major factors in the FRS and ASCVD pooled cohort equations score, are closely related to lifestyle changes. Similar to previous studies, the effect of a single coronary computed tomography scan and the results of counseling decreased over time, and the differences that were observed after 2 years disappeared after 5 years.

**Limitations**

This study has some limitations. First, the follow-up period was 5 years, and the risk scores were not observed for a longer period (eg, 10 years, as CVDs can last for >10 years). A follow-up study for identifying a risk score that is suitable for CVD prediction over longer periods can be conducted in the future. Second, as this was a single-center study, some of the outcomes may not be generalizable. Multicenter studies that use OHDSI data networks can provide more generalizable evidence. Third, this study included patients who visited the health promotion center multiple times; those who did not undergo CCTA at the first visit but underwent CCTA during subsequent visits were included in the comparison group. Therefore, the differences between the groups might have been attenuated. This can be avoided by conducting a prospective cohort study. Lastly, observational research that uses EHR data has the limitation that it cannot fully capture the entirety of a patient's health information [21]. This study converted EHR data into common data model data, and it has the same limitation. If the participants of this study underwent examinations and treatments outside of the hospital, there was a disadvantage that the records for these procedures were not recorded in the database. Additionally, with regard to drugs, the SNUBH common data model converted data on prescription drugs for outpatients and administration drugs for inpatients. Thus, it was not known whether the drugs ordered for the outpatients were taken on time by the patients. As such, selection bias may have occurred due to information not being recorded in the database. Although it is possible to reduce channeling bias through large-scale propensity score matching, which we used in this study, there may still be the limitation that such matching cannot reduce selection bias [22].

**Discussion**

**Principal Results**

From our population-level estimation study, which compared the CVD HRs of a health checkup group that was undergoing CCTA with those of a group that was not undergoing CCTA over 5 years, although some benefits were observed at 2 years, we found no significant difference (calibration $P=.37$) in the final risk of CVD events between the two groups. It seems that CCTA has no beneficial effect on CVD prevention for long periods of time.

Communication about medical examinations and examination results through counseling has been reported to improve health indicators, such as CVD risk. In the Korean national health insurance service screening program, the group that underwent cardiovascular health screening for 40-year-olds had higher rates of new hypertension, diabetes, and hyperlipidemia, whereas the incidence of CVD mortality, all-cause mortality, and major adverse cardiovascular events was lower [15]. Per the results of an analysis of the same data, the group that received counseling after the health checkup had higher motivation stages of health behavior change than those of the group that received only the checkup [16]. The smoking cessation rate was higher after 2 years when compared to that of the group who received only the checkup [17]. Engberg et al [18] reported that cardiovascular risk scores, BMIs, and serum cholesterol levels were lower in the intervention groups than those in the control group after 5 years’ worth of health screenings and consultations.

In existing studies that require lifestyle modifications, such as modifications for obesity, smoking cessation, and substance abuse, the effects of 1-time interventions or short-term interventions, interviews, and counseling tend to weaken over time. In a study that used the motivational interview technique for people with substance abuse issues, the positive effect observed at 3 months disappeared at 12 months [19], and in another study, the effect of smoking cessation treatment continued for 10 weeks and gradually slowed down at 3, 6, and 12 months [20].

Our study compared patients who did or did not undergo additional coronary computed tomography. Both groups underwent the same levels of examination and counseling, which were conducted by the cardiovascular health screening program of the national service in 1 hospital.

Smoking status, blood pressure, and blood lipid concentration, which are major factors in the FRS and ASCVD pooled cohort equations score, are closely related to lifestyle changes. Similar to previous studies, the effect of a single coronary computed tomography scan and the results of counseling decreased over time, and the differences that were observed after 2 years disappeared after 5 years.

**Limitations**

This study has some limitations. First, the follow-up period was 5 years, and the risk scores were not observed for a longer period (eg, 10 years, as CVDs can last for >10 years). A follow-up study for identifying a risk score that is suitable for CVD prediction over longer periods can be conducted in the future. Second, as this was a single-center study, some of the outcomes may not be generalizable. Multicenter studies that use OHDSI data networks can provide more generalizable evidence. Third, this study included patients who visited the health promotion center multiple times; those who did not undergo CCTA at the first visit but underwent CCTA during subsequent visits were included in the comparison group. Therefore, the differences between the groups might have been attenuated. This can be avoided by conducting a prospective cohort study. Lastly, observational research that uses EHR data has the limitation that it cannot fully capture the entirety of a patient's health information [21]. This study converted EHR data into common data model data, and it has the same limitation. If the participants of this study underwent examinations and treatments outside of the hospital, there was a disadvantage that the records for these procedures were not recorded in the database. Additionally, with regard to drugs, the SNUBH common data model converted data on prescription drugs for outpatients and administration drugs for inpatients. Thus, it was not known whether the drugs ordered for the outpatients were taken on time by the patients. As such, selection bias may have occurred due to information not being recorded in the database. Although it is possible to reduce channeling bias through large-scale propensity score matching, which we used in this study, there may still be the limitation that such matching cannot reduce selection bias [22].

**Table 3.** Changes in the differences in American College of Cardiology/American Heart Association (ACC/AHA) atherosclerotic cardiovascular disease (ASCVD) risk scores and Framingham risk score (FRSs) from baseline at 2 and 5 years.

<table>
<thead>
<tr>
<th></th>
<th>CCTA group</th>
<th></th>
<th>Non-CCTA group</th>
<th></th>
<th>$P$ value$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Differences in risk scores from baseline at 2 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACC/AHA ASCVD risk scores</td>
<td>1330</td>
<td>0.17 (−0.16 to 1.08)</td>
<td>1691</td>
<td>0.23 (−0.10 to 1.30)</td>
<td>.03</td>
</tr>
<tr>
<td>FRSs</td>
<td>1330</td>
<td>0.39 (−0.80 to 1.96)</td>
<td>1691</td>
<td>0.60 (−0.69 to 2.26)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Differences in risk scores from baseline at 5 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACC/AHA ASCVD risk scores</td>
<td>1232</td>
<td>1.10 (0.08 to 1.57)</td>
<td>1372</td>
<td>1.06 (0 to 2.79)</td>
<td>.97</td>
</tr>
<tr>
<td>FRSs</td>
<td>1232</td>
<td>1.66 (0.04 to 3.92)</td>
<td>1372</td>
<td>1.61 (0.09 to 4.11)</td>
<td>.85</td>
</tr>
</tbody>
</table>

$^a$CCTA: coronary artery computed tomography angiography.

$^b$Wilcoxon rank-sum test $P$ value.
Comparison With Prior Work
Waugh et al [23] conducted a systematic review and meta-analysis of 5 studies and reported that computed tomography has no benefits as a screening tool for the potential onset of CVDs. However, a closer review revealed that all 5 included studies were inappropriate in terms of their findings about the prophylactic benefits of CCTA. All of these studies investigated the association between coronary artery calcium (CAC) and the onset of CVDs or death after a specific follow-up period in patients who underwent CCTA screening. They used a short follow-up period and analyzed the results in the context of the presence of CAC as opposed to CCTA findings. Therefore, the conclusion of the meta-analysis by Waugh et al [23]—CCTA screening is not effective—was based on the finding that the risk of heart disease was not elevated in people undergoing a CAC assessment via CCTA, as opposed to an assessment of the prophylactic benefits of CCTA itself. Further, since the measurement of CAC is regarded as a reliable method for CVD risk assessment, a study claimed that CCTA should be introduced for the screening of asymptomatic individuals [24]. However, other studies claim that CCTA is cost-ineffective, although these admit that CAC, when observed via CCTA, is a better predictor of CVD than the FRS [25]. We supplemented these studies by comparing groups that underwent CCTA with those that did not undergo CCTA.

McEvoy et al [26] examined the differences in the incidence of coronary artery disease between CCTA and comparison groups after a fixed follow-up period. The authors matched the propensity scores of 1000 individuals who underwent CCTA for a health checkup with those of 1000 individuals who did not undergo CCTA (ie, the comparison group) and compared the incidence of coronary artery disease at the 90-day and 18-month follow-ups. The study reported that CCTA-based screening was significantly associated with an increased rate of invasive tests and medication use but was not associated with the incidence of coronary artery disease, concluding that CCTA is not recommended for screening purposes. However, the study was limited by the small number of cases and the short follow-up periods.

Our study presents reliable evidence about CCTA, which was obtained by performing large-scale propensity score matching and using EHR and health checkup questionnaire responses from OMOP-CDM data. We studied a large study sample over a longer study period than those used by previous studies. Although past studies used either 90-day follow-ups or 18-month follow-ups, we observed the patients from 60 days after the index date to 5 years after the index date to analyze the CVD HRs in relation to CCTA. Moreover, while previous studies had approximately 1000 patients in both the target group and comparison group, we included 8787 patients in each group. The data were also standardized, which enabled us to perform an efficient analysis across organizations and use the same analysis codes. Future studies can investigate the effects of CCTA and CVD in larger populations over long follow-up periods, in collaboration with organizations that convert health questionnaire data into the common data model format.

We also stratified the population into high-risk and low-risk groups based on the ASCVD risk score and FRS. Even in the high-risk group, CCTA screening did not have a significant effect (ASCVD risk score: calibration $P$=.39; FRS: calibration $P$=.13) on the prevention of CVD.

Based on the changes in risk scores, a significant difference was observed between the CCTA and comparison groups after 2 years (change in ASCVD risk scores: $P$=.03; change in FRSs: $P$=.02). However, this difference was not significant after 5 years (change in ASCVD risk score: $P$=.92; change in FRSs: $P$=.85). We speculate that patients are motivated to manage their risk score factors for a brief period immediately after the CCTA test; however, the significance decreases over long periods.

Conclusions
Through a retrospective cohort study that was conducted over a 5-year period, we found that CCTA had no significant preventive effect on future CVDs. We also demonstrated the potential of converting health checkup data into OMOP-CDM data and integrating such data into common data model–based EHR data for research targeting the health checkup population. Although we examined the outcomes of CVDs after CCTA, future studies could examine patients’ health behaviors following CCTA. It is expected that the use of common data model data will be expanded to multicenter studies.

Acknowledgments
This work was supported by the Technology Innovation Program (grant 20004927 for “Upgrade of CDM based Distributed Biohealth Data Platform and Development of Verification Technology”), which is funded by the Ministry of Trade, Industry & Energy (Korea).

Data Availability
Common data model data are designed to support a distributed research network. Thus, access to the data is restricted on internal private networks, and the data are not publicly available.

Authors’ Contributions
WKB designed this study. JC drafted the manuscript and performed the data analyses. SK, BK, HB, and WS reviewed the data extraction and study design. WKB, JC, and SY inspected and revised the manuscript. SY supervised this study. All authors have read and approved the final manuscript.
Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.

References


Abbreviations

ACC/AHA: American College of Cardiology/American Heart Association
ASCVD: atherosclerotic cardiovascular disease
CAC: coronary artery calcium
CCTA: coronary artery computed tomography angiography
CHA2DS2-VASc: Congestive Heart Failure, Hypertension, Age≥75 (Doubled), Diabetes, Stroke (Doubled), Vascular Disease, Age 65 to 74, and Sex Category (Female)
CHADS2: Congestive Heart Failure, Hypertension, Age, Diabetes, Previous Stroke/Transient Ischemic Attack (2 points)
CVD: cardio-cerebrovascular disease
EHRR: electronic health record
FRS: Framingham risk score
HR: hazard ratio
ICD-10: International Classification of Diseases, 10th Revision
IRB: Institutional Review Board
OHDSI: Observational Health Data Sciences and Informatics
OMOP-CDM: Observational Medical Outcomes Partnership Common Data Model
SNUBH: Seoul National University Bundang Hospital

Please cite as:
Bae WK, Cho J, Kim S, Kim B, Baek H, Song W, Yoo S
Coronary Artery Computed Tomography Angiography for Preventing Cardio-Cerebrovascular Disease: Observational Cohort Study Using the Observational Health Data Sciences and Informatics’ Common Data Model
JMIR Med Inform 2022;10(10):e41503
URL: https://medinform.jmir.org/2022/10/e41503
PMID:36227638
A Recurrent Neural Network Model for Predicting Activated Partial Thromboplastin Time After Treatment With Heparin: Retrospective Study

Sebastian Daniel Boie¹, PhD; Lilian Jo Engelhardt¹,², MD; Nicolas Coenen², MD; Niklas Giesa¹, MSc; Kerstin Rubarth¹,³, MSc; Mario Menk¹,², MD; Felix Balzer¹, PhD, MD

¹Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Medical Informatics, Berlin, Germany
²Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Anesthesiology and Intensive Care Medicine, Berlin, Germany
³Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Biometry and Clinical Epidemiology, Berlin, Germany

Corresponding Author:
Sebastian Daniel Boie, PhD
Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin
Institute of Medical Informatics
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30 450580877
Email: Sebastian-Daniel.Boie@charite.de

Abstract

Background: Anticoagulation therapy with heparin is a frequent treatment in intensive care units and is monitored by activated partial thromboplastin clotting time (aPTT). It has been demonstrated that reaching an established anticoagulation target within 24 hours is associated with favorable outcomes. However, patients respond to heparin differently and reaching the anticoagulation target can be challenging. Machine learning algorithms may potentially support clinicians with improved dosing recommendations.

Objective: This study evaluates a range of machine learning algorithms on their capability of predicting the patients’ response to heparin treatment. In this analysis, we apply, for the first time, a model that considers time series.

Methods: We extracted patient demographics, laboratory values, dialysis and extracorporeal membrane oxygenation treatments, and scores from the hospital information system. We predicted the numerical values of aPTT laboratory values 24 hours after continuous heparin infusion and evaluated 7 different machine learning models. The best-performing model was compared to recently published models on a classification task. We considered all data before and within the first 12 hours of continuous heparin infusion as features and predicted the aPTT value after 24 hours.

Results: The distribution of aPTT in our cohort of 5926 hospital admissions was highly skewed. Most patients showed aPTT values below 75 s, while some outliers showed much higher aPTT values. A recurrent neural network that consumes a time series of features showed the highest performance on the test set.

Conclusions: A recurrent neural network that uses time series of features instead of only static and aggregated features showed the highest performance in predicting aPTT after heparin treatment.

(JMIR Med Inform 2022;10(10):e39187) doi:10.2196/39187

KEYWORDS
machine learning; health care; recurrent neural network; heparin; activated partial thromboplastin time (aPTT); deep learning; ICU; critical care
**Introduction**

Thromboembolic complications are associated with increased mortality [1,2]. Risk factors for deep venous thrombosis and pulmonary embolism include, for example, immobility, malignancy, higher age, and a history of thromboembolism [3,4]. Anticoagulation by drugs is applied either prophylactically to prevent thromboembolism [5] or therapeutically to treat existing thromboembolic complications, which reduces mortality [6].

In perioperative normal care wards, prophylactic and therapeutic anticoagulation is frequently performed subcutaneously by low–molecular weight heparins [5]. In the perioperative setting, prophylactic anticoagulation is indicated in patients with intermediate or high risk for thromboembolism. This includes, for example, most trauma surgeries, elective orthopedic surgeries with consecutive immobility of the lower limbs, and major abdominal or thoracic surgery, particularly in the presence of malignant and inflammatory processes [5].

In critical illness, the risk for venous thromboembolism is increased in almost all patients due to the combination of general risk factors related to chronic disease and intensive care unit (ICU)–associated risk factors, including sedation, immobility, or central venous catheters [7]. In intensive care, prophylactic or therapeutic anticoagulation is regularly applied intravenously by continuous unfractionated heparin, particularly during renal failure or hemodynamic instability [8]. The short half-life of the anticoagulant and the possibility of antagonizing heparin with protamine are advantages of unfractionated heparin in these vulnerable patients [9]. However, poor controllability is an issue. Consequently, overdosing with hemorrhagic or underdosaging with thrombotic complications may occur [10]. Hence, therapeutic unfractionated heparin application requires monitoring. The dosing of unfractionated heparin is performed by determination of activated partial thromboplastin time (aPTT) in patients’ blood [11]. Based on older studies, the pursued aPTT target is approximately a 1.5 to 2.5-times prolongation of the reference clotting time [11-13] although individual targets are usually defined. Achieving the aPTT target within 24 hours has been associated with increased survival in patients with pulmonary embolism [6]. However, due to patient- and disease-related variations, achieving the aPTT target within 24 hours is challenging.

Nowadays, big data sets are generated by digital patient data management systems in ICU routine. Machine learning (ML) approaches that include individual information from large data sets may help to predict aPTT at an earlier stage than can routine blood sampling. Previous results of applying ML to predict aPTT show great promise [14-17]. Some authors [16,17] consider the numerical value of aPTT and consequently the prediction of aPTT as a regression task. We prefer the prediction of the numerical value since it makes no assumption of the aPTT target range. However, most recent literature on similar-sized data sets consider aPTT after heparin treatment as a multiclass prediction with 3 distinct ranges (subtherapeutic, therapeutic, or supratherapeutic) [14,15,18].

In previous model comparison studies [15,16,18], it has been demonstrated that artificial neural networks show the highest performance on aPTT prediction tasks.

Recently, a systematic review of ML approaches on predicting aPTT after heparin administration highlighted that still multiple innovations are required before ML-assisted heparin dosing is ready for clinical practice [19].

We compared multiple ML models on our patient cohort and are, to our knowledge, the first to apply a recurrent neural network model that takes the dynamics of variables in the form of time series into account. At the outset of the study, we specified inclusion criteria that resulted in 5926 distinct hospital admissions. On this cohort, we trained and evaluated multiple ML models on the aPTT prediction task. To allow comparison of the recurrent neural network model with previously published models [14,15,18], we subsequently used our model in a classification setup.

The aim of this analysis is to evaluate whether ML models can accurately predict subsequent aPTT measurements well (12 hours) in advance. In the future, data-driven approaches could help clinicians to adjust heparin dosing to improve time in the target range aPTT after 24 hours.

**Methods**

**Data Selection Criteria**

The database system for surgical and intensive care patients at Charité – Universitätsmedizin Berlin (Charité) was first adopted in 2012 and over time rolled out to all ICUs. Since we extracted data in November 2021, we considered a time period from 2012 to October 31, 2021. We selected patients and hospital admissions that satisfied the following inclusion criteria: at least 18 years old at the beginning of treatment, received a minimum of 1000 IU of heparin, received some of the heparin as continuous infusion, had at least a single aPTT measurement after 12 hours and before 36 hours after the intravenous treatment commenced, and had weight and height documented (within reasonable limits: height between 25 cm to 250 cm, weight between 3 kg to 300 kg).

**Ethics Approval**

Ethics approval for this study was obtained by the Charité ethics committee (vote #EA4/241/21).

**Feature Selection and Prediction Targets**

We extracted patient characteristics (age, gender, height, weight), laboratory values (aPTT, bilirubin, C-reactive protein, creatinine, quick value, platelet count), whether patients received dialysis or a form of extracorporeal membrane oxygenation (ECMO) treatment, and routinely collected scores (therapeutic intervention scoring system 10 [TISS-10], simplified acute physiology score [SAPS-II], sequential organ failure assessment [SOFA], acute physiology and chronic health evaluation II [APACHE II]) from the hospital information system. Furthermore, we extracted the time of the start and end of each heparin dosing, concentration, and administration rate. Heparin can be administered as a bolus or as a continuous infusion. All
data were restricted to the time period 7 days prior to treatment to 36 hours after treatment started.

Our goal was to predict the aPTT 24 hours after initiation of continuous heparin treatment. However, not all patients had a laboratory measurement exactly 24 hours after the treatment with heparin. Thus, any aPTT measurement between 12 to 36 hours after heparin treatment began was accepted as the prediction target. In case multiple values were recorded between 12 hours and 36 hours, we chose the value that was closest to 24 hours after continuous treatment started. Consequently, only values that were taken before or within 12 hours after continuous heparin treatment commenced were available as features for the model (including any aPTT measurement in that time frame).

Hospital stays were left aligned, and the start of the continuous intravenous heparin delivery corresponded to time zero.

Handling of Missing Data

The data we used for our study were collected during routine care and were not of uniform quality across all hospital admissions. A typical problem when using retrospective data for ML is missing observations [20-22]. This problem is exacerbated for the recurrent neural network, as it expects an input for every feature every 2 hours.

The static values of gender, age, height, and weight had no missing values and were replicated for every timestamp. The one-hot-encoded variables, including ECMO treatment, dialysis, bolus delivery of heparin, and continuous delivery of heparin, were set to 0 if no other value was recorded for a given timestamp. Other features (eg, laboratory measurements and scores) were filled in a 2-step process as follows: (1) If a previous value was recorded within 7 days prior to continuous heparin treatment, those values were forward filled; (2) Any still missing values were replaced by the mean across the training population.

Only using the above 2-step procedure discards information about which measurement is from the patient at the given timestamp. Since it has been shown that the missing pattern can be informative [23], we included an “indicator” variable for each variable filled in the 2-step process that is 1 if the value was measured at the given timestamp and 0 if it was imputed.

Together with the indicator variables, each model sees 35 different input variables.

The recurrent neural network, thus, may see time series between $t = -168$ (7 days prior to continuous heparin delivery) to $t = 12$. In general, however, patients’ time series are not of the same length.

Models and Variable Encoding

The input data consisted of numerical and categorical variables. Categorical variables (gender, ECMO treatment, dialysis treatment, continuous heparin administration, bolus heparin administration) were one-hot encoded. Each option for a categorical variable resulted in 1 input dimension that could either be 1 or 0. One-hot–encoded variables were not further scaled and were directly used as input features.

Other numerical variables were standardized before being fed into the model. Mean and SD were estimated only on the training data set.

We compared 6 models that take a single value per feature and 1 model that takes the entire time series of features. Some features were constant over the course of treatment (age, gender, height, and weight), while the other features changed frequently. Models that take a single value per feature received the last-observed value before the 12-hour cutoff. The recurrent neural network received time series, resampled to 2-hour intervals, for each feature. If multiple measurements were taken within 2 hours, those values were replaced by the mean over this 2-hour window. Static variables were repeated for each timestamp. The prediction target (a single aPTT measurement) is log-transformed during model training. The log transformation is discussed in the Results section. All model parameters are optimized on the mean-squared-error (MSE) loss function. Additionally, we evaluated the mean absolute error and the explained variance for each model.

The 6 regression models were linear regression, elastic net, generalized linear model, support vector machine regression (SVR), K-nearest neighbor regression (KNN), and regression trees. We optimized hyperparameters using a grid search with 5-fold cross-validation. For the cross-validation, training and validation data were combined. The hyperparameter grids are shown in Table 1.

The models, cross-validation, and the grid search routine were from the scikit-learn package [24] and implemented in Python (The Python Software Foundation).
**Table 1.** Hyperparameters for each static model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>None</td>
</tr>
<tr>
<td>Elastic net</td>
<td>$\alpha = (10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 2, 3)$ $L_1$-ratio = (0, 0.1, … 1.0)</td>
</tr>
<tr>
<td>GLM$^a$</td>
<td>$\alpha = (10^{-2}, 10^{-1}, 1, 2, 3)$ Power = (0, 1, 2, 3)</td>
</tr>
<tr>
<td>SVR$^b$</td>
<td>Kernel = (“linear,” “poly,” “rbf,” “sigmoid”) Degree = (2, 3, 4, 5, 6)</td>
</tr>
<tr>
<td>KNN$^c$</td>
<td>$K = (2, 3, 4, 5, 6, 7, 8, 9, 10)$ $\text{Weights} = (\text{&quot;uniform,&quot; &quot;distance&quot;})$</td>
</tr>
<tr>
<td>Regression trees</td>
<td>$\text{Max_depth} = (2, 3, 4, 5, \text{unlimited})$ $\text{Min_samples_split} = (2, 3, 4, 5, 6)$ $\text{Min_samples_leaf} = (1, 2, 3, 4, 5)$</td>
</tr>
</tbody>
</table>

$^a$GLM: generalized linear model.  
$^b$SVR: support vector machine regression.  
$^c$KNN: K-nearest neighbor regression.

**Recurrent Neural Network Model**

This model consists of a gated recurrent unit (GRU), which can process a time series of arbitrary length and a fully connected network that uses the output of the GRU as input. Since we are only interested in predicting a single value, only the last output of the GRU is fed into a 3-layer fully connected model. No activation function is used between the output of the GRU and the first fully connected layer. The outputs of the 2 fully connected layers have rectified linear unit activation functions [25], and the final layer has no activation function. A schematic overview can be seen in Figure 1.

**Figure 1.** Schematic overview of input features, recurrent neural network, and feedforward network. GRU: gated recurrent unit.

As for the previously described models, the recurrent neural network was optimized on the MSE. For experiments with the recurrent neural network, weights were optimized on the training set, and the results between experiments were compared on the validation set. We used the Adam optimizer with L2 penalty [26]. For each experiment, we chose weights with the lowest error on the validation set, which may occur before the maximum number of epochs are reached.

This model is significantly more costly to train compared to “static” models. Therefore, we did not perform a systematic hyperparameter optimization but ran several experiments with different hyperparameters and handpicked the best set of hyperparameters, which are shown in the Results section. Hyperparameters for the GRU submodel are hidden size ($n=1, 2, 3, \ldots$), bidirectional connection (True, False), and the number of layers ($n=1, 2, 3, \ldots$).

The 3-layered fully connected submodel had the number of neurons in each layer as 3 hyperparameters. Hyperparameters related to the training are the learning rate, L2 penalty, and the maximum number of epochs.

Patients have a different number of inputs per feature, since they receive their continuous heparin treatment at different times during their hospital stay. Thus, for training, we are limited to a batch size of 1 but accumulate multiple batches before weights are updated. To combat overfitting, we used an L2 penalty on the weights in the fully connected part of the model and chose weights on the epoch with the highest performance on the validation set.

All models and training scripts are available on github [27].

**Classification Models**

To phrase aPTT prediction as a classification task, we used the 3 ranges first introduced by Ghassemi et al [14] of
subtherapeutic for values below 60 s, therapeutic for values between 60 s to 100 s, and supratherapeutic for values above 100 s for the aPTT measurements. We compared our GRU model to the logistic regression model from Ghassemi et al [14] and the feedforward neural networks models by Su et al [15] and Li et al [18]. All parameters were taken from the reference literature for the respective model. For the feedforward networks from Su et al [15] and Li et al [18], we used cross-entropy [28] as a loss function with early stopping since the loss functions are not mentioned in the references.

The 3 classification models are retrained on the training split and receive the last value of each feature before the 12-hour cutoff in the same manner as the “static” regression models. The GRU is not retrained on the classification task, but the numeric predictions are binned into the 3 ranges post hoc. We evaluated the models on macroaveraged precision, macroaveraged recall, macroaveraged F1-score, and accuracy [29].

**Results**

**Patient Cohort**

A flow diagram of consecutively applied filter criteria (specified in the methods section) to the entire patient cohort is shown in **Figure 2**. The selection criteria resulted in 5926 hospital admissions from a total of 5742 unique patients. Given that fewer than 4% of admissions occurred for previously admitted patients, we considered hospital admissions to be independent events. Basic patient characteristics and missing values are documented in **Table 2**.

Before model training or parameter estimation for mean and SD were performed, the admissions were split into training (n=3800), validation (n=945), and test (n=1181) samples. We ensured that different admissions by the same patient were in the same fold.

**Figure 2.** Flow diagram of unique patients and admissions that satisfy the specified inclusion criteria. aPTT: activated partial thromboplastin time; IV: intravenous line.
Table 2. Basic characteristics of the study cohort. The third column indicates how many patients do not have a single measurement during the hospital admission.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value (N=5926)</th>
<th>Patients missing for entire stay, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), median (IQR)</td>
<td>70.62 (60.95-77.74)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1910 (32)</td>
<td>N/A^a</td>
</tr>
<tr>
<td>Male</td>
<td>4016 (68)</td>
<td>N/A</td>
</tr>
<tr>
<td>Height (cm), median (IQR)</td>
<td>172 (164-178)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Weight (kg), median (IQR)</td>
<td>77 (66-90)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>SOFA^b, median (IQR)</td>
<td>5 (2-8)</td>
<td>442 (7.46)</td>
</tr>
<tr>
<td>SAPS II^c, median (IQR)</td>
<td>36 (27-47)</td>
<td>449 (7.58)</td>
</tr>
<tr>
<td>APACHE II^d, median (IQR)</td>
<td>17 (12-23)</td>
<td>525 (8.86)</td>
</tr>
<tr>
<td>TISS-10^e, median (IQR)</td>
<td>10 (5-15)</td>
<td>5755 (97.11)</td>
</tr>
<tr>
<td>Dialysis, n (%)</td>
<td>449 (7.57)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>ECMO^f, n (%)</td>
<td>76 (1.28)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>aPTT^g (s), median (IQR)</td>
<td>42.6 (36.1-54.6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Bilirubin (mg/dL), median (IQR)</td>
<td>0.6 (0.35-1.24)</td>
<td>2529 (42.69)</td>
</tr>
<tr>
<td>CRP^h (mg/L), median (IQR)</td>
<td>56.2 (18.6-118.8)</td>
<td>1782 (30.07)</td>
</tr>
<tr>
<td>Gfr^i (count), median (IQR)</td>
<td>67 (39-90)</td>
<td>71 (1.20)</td>
</tr>
<tr>
<td>Creatinine (mg/dL), median (IQR)</td>
<td>1.01 (0.74-1.56)</td>
<td>32 (0.54)</td>
</tr>
<tr>
<td>Quick value (%), median (IQR)</td>
<td>76 (64-87)</td>
<td>17 (0.29)</td>
</tr>
<tr>
<td>Platelet count (per nL), median (IQR)</td>
<td>204 (139-292)</td>
<td>19 (0.32)</td>
</tr>
<tr>
<td>Total heparin administered (IU), median (IQR)</td>
<td>32398 (9500-90000)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

^aN/A: not applicable.
^bSOFA: sequential organ failure assessment.
^cSAPS II: simplified acute physiology score II.
^dAPACHE II: acute physiology and chronic health evaluation II.
^eTISS-10: therapeutic intervention scoring system 10.
^fECMO: extracorporeal membrane oxygenation.
^g aPTT: activated partial thromboplastin time.
^hCRP: C-reactive protein.
^iGfr: glomerular filtration rate.

**Distribution of aPTT Values**

A histogram of measured aPTT before and after treatment is shown in Figure 3. In our cohort, both aPTT distributions before and after heparin treatment are narrowly peaked with a heavy tail. Values above 100 s occur very rarely. Small peaks are visible at 240 s where the laboratory reports some values as >240 s, which is mapped to 240 s.
The effect of heparin treatment on the entire cohort is clearly seen by the shift of the distribution. The difference in means is 8.64 s (95% CI 7.72-9.56; \(P<.001\)). The first 4 moments of the distribution of aPTT at \(t=0\) and at \(t=24\) are documented in Table 3. The mean aPTT value is higher after continuous heparin delivery compared to before treatment. Skew and kurtosis (while smaller after treatment) quantifiably indicate that the aPTT distribution is not symmetric and has a heavy tail. This fact makes the prediction of aPTT challenging. To make the learning task easier for our models, we log-transform the target variable to reduce skew and kurtosis. In effect, this makes "rare" events in the original distribution easier to predict.

The distribution that we observed in the Charité cohort contrasts with the aPTT values that are documented by other authors. Su et al [15] and Ghassemi et al [14] base their modeling studies on the Medical Information Mart for Intensive Care (MIMIC) II/III and eICU databases. The distribution of aPTT on the eICU database [15] is more heavy tailed than is the MIMIC cohort, however, less so than is our cohort. The 3 treatment categories reported in those works are indicated as shaded regions in Figure 3b. However, we do not classify our cohort into these categories but treat the prediction of aPTT after treatment as a regression problem.

### Table 3. Statistical description of the binned distribution of aPTT values before continuous heparin treatment \((t=0)\), 24 hours after continuous treatment commenced \((t=24)\), and the log-transformed distribution after 24 hours.

<table>
<thead>
<tr>
<th></th>
<th>(aPTT^3) ((t=0))</th>
<th>(aPTT) ((t=24))</th>
<th>Log ((aPTT \ [t=24]))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations, n</td>
<td>4850</td>
<td>5926</td>
<td>5926</td>
</tr>
<tr>
<td>Mean</td>
<td>40.64</td>
<td>49.28</td>
<td>3.83</td>
</tr>
<tr>
<td>Variance</td>
<td>561.55</td>
<td>608.19</td>
<td>0.11</td>
</tr>
<tr>
<td>Skew</td>
<td>6.11</td>
<td>4.74</td>
<td>1.91</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>42.93</td>
<td>26.71</td>
<td>5.37</td>
</tr>
</tbody>
</table>

\(aPTT\): activated partial thromboplastin time.

### Model Comparisons

In this section, the results of comparing 7 different models on the prediction of aPTT (see Table 4) are shown. Models 1-6 received only the last-measured values of each input feature before the 12-hour cutoff. We optimized hyperparameters for each model using a grid search and 5-fold cross-validation. The reported results are based on the test data that was not included in the 5 folds. A full description of the used grids appears in the Methods section. The best parameters for Models 1-6 are documented in Multimedia Appendix 1.

Model 7 (recurrent neural network) consumes the entire time series, resampled to 2-hour timestamps, for each input feature. We experimented also with resampling to 1-hour time steps and 4-hour time steps and found that the performance was similar (see Multimedia Appendix 1 for numerical results).

It is the most complex model in the comparison and ingests data from up to 7 days before continuous treatment to 12 hours after continuous treatment is administered. A systematic hyperparameter optimization for Model 7 was not performed;
hence, we are underestimating the performance of the recurrent neural network in comparison to other models. However, the recurrent neural network model achieved the highest score on the explained variance and MSE metrics. It ranked second to the SVR model on the mean absolute error (which penalizes outliers less than does the MSE). The SVR models ranked second to the recurrent neural network model on explained variance and MSE. CIs were obtained by taking 1000 random samples of the same size as the test set, with replacement. Given that the distribution had a small number of large outliers, which had a significant effect on the quantity of interest, the CIs are wide.

<table>
<thead>
<tr>
<th>Model</th>
<th>Explained variance</th>
<th>MSE&lt;sup&gt;a&lt;/sup&gt;</th>
<th>MAE&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Linear regression, test set value, (95% CI)</td>
<td>0.163 (0.115-0.211)</td>
<td>0.487 (0.425-0.556)</td>
<td>0.474 (0.45-0.497)</td>
</tr>
<tr>
<td>2  Elastic net regression</td>
<td>0.168 (0.124-0.214)</td>
<td>0.484 (0.433-0.554)</td>
<td>0.474 (0.453-0.497)</td>
</tr>
<tr>
<td>3  GLM&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.169 (0.121-0.21)</td>
<td>0.484 (0.422-0.556)</td>
<td>0.473 (0.450-0.5)</td>
</tr>
<tr>
<td>4  Support vector regression</td>
<td>0.203 (0.161-0.244)</td>
<td>0.476 (0.406-0.554)</td>
<td>0.442 (0.418-0.469)</td>
</tr>
<tr>
<td>5  Nearest neighbors</td>
<td>0.101 (0.055-0.140)</td>
<td>0.529 (0.460-0.597)</td>
<td>0.502 (0.477-0.528)</td>
</tr>
<tr>
<td>6  Decision tree regression</td>
<td>0.154 (0.108-0.198)</td>
<td>0.492 (0.427-0.563)</td>
<td>0.471 (0.447-0.495)</td>
</tr>
<tr>
<td>7  Recurrent NN&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.21 (0.165-0.254)</td>
<td>0.459 (0.4-0.523)</td>
<td>0.454 (0.432-0.477)</td>
</tr>
</tbody>
</table>

<sup>a</sup>MSE: mean-squared error.  
<sup>b</sup>MAE: mean absolute error.  
<sup>c</sup>GLM: generalized linear model.  
<sup>d</sup>NN: neural network.

**Prediction of aPTT by the Recurrent Neural Network Model**

In this section we present the results of the recurrent neural network model and compare predictions with measured values on the test set. Multiple experiments with the model were performed, and the best handpicked parameters are shown in Table 5.

Predictions and measurements are shown in Figure 4. The distributions of aPTT values in the test data alone show a similar distribution as the aPTT values over the entire data set (cf Figure 3 and Figure 4 right panel). The histogram of predictions of the recurrent neural network model has a similar shape (cf Figure 4 top panel and Figure 4 right panel).

Direct comparisons between predictions and measurements can be seen in the center of Figure 4. The model can predict the majority of aPTT values very well. Although some outliers are predicted accurately, there are a few outliers above 150 s where predictions fall below 75 s. Likewise, some predicted outliers do not manifest as actual outliers.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate</td>
<td>$10^{-3}$</td>
</tr>
<tr>
<td>Layers</td>
<td>Single GRU&lt;sup&gt;a&lt;/sup&gt; layer; 3 feedforward layers with 10, 5, and 1 output neurons, respectively</td>
</tr>
<tr>
<td>Hidden size (GRU)</td>
<td>5</td>
</tr>
<tr>
<td>Bidirectional</td>
<td>True</td>
</tr>
<tr>
<td>Accumulate gradient batches</td>
<td>16</td>
</tr>
<tr>
<td>L2 penalty on all weights</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<sup>a</sup>GRU: gated recurrent unit.
Figure 4. Predictions versus measurements. The figure shows predicted (abcissa) and measured aPTT (ordinate) after 24 hours in the central panel. Only predictions on the test set are shown. The dashed diagonal line indicates a perfect match between prediction and measurement. Above and to the right are binned distributions of all predictions and measurements, respectively. aPTT: activated partial thromboplastin time.

Comparison With Classification Models

In the previous sections, we have seen that the recurrent neural network shows the highest performance on the regression task. However, it is also apparent that not all predictions are accurate. To understand whether improvements needed to occur on the models or on data quality aspects, we rephrased the problem as a classification task to be able to compare the performance of the trained model with the 3 most recently published classification models [14,15,18]. Each of the 3 models was trained on our data set (details in the Methods section).

Our recurrent neural network scored the highest performance in recall and $F_1$-score. The simplest model (logistic regression by Ghassemi et al [14]) had the highest precision, and the feedforward neural network by Li et al [18] had the highest accuracy (see Table 6 for results). No single model outperformed the others on all 4 metrics, and the appropriate model may be chosen depending on which metric is considered most relevant.

The fact that the best-published models show a comparable performance indicate that significant improvements require a closer monitoring of patients, additional tests, and improved data quality.

Table 6. Comparison of different models when formulating activated partial thromboplastin time prediction as a classification task. For each metric, a higher score is better.

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$-score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRU</td>
<td>0.411</td>
<td>0.396</td>
<td>0.398</td>
<td>0.829</td>
</tr>
<tr>
<td>Ghassemi [14]</td>
<td>0.707</td>
<td>0.357</td>
<td>0.356</td>
<td>0.825</td>
</tr>
<tr>
<td>Su [15]</td>
<td>0.357</td>
<td>0.338</td>
<td>0.316</td>
<td>0.834</td>
</tr>
<tr>
<td>Li [18]</td>
<td>0.430</td>
<td>0.350</td>
<td>0.338</td>
<td>0.838</td>
</tr>
</tbody>
</table>

$^a$GRU: gated recurrent unit.
Discussion

Principal Findings

In this study, we analyzed and predicted the effect of heparin treatment on a cohort of 5742 patients and 5926 hospital admissions 24 hours after continuous application. A statistically significant shift of aPTT measurements compared to the beginning of the treatment was observed. Most patients’ aPTT measurements were within 35 s to 75 s; however, some patients showed much higher aPTT values, leading to a challenging prediction problem with a long-tailed distribution. We demonstrated that ML models can aid in predicting the aPTT values 12 hours in advance. Additionally, we have shown that using the time series of variables improves predictive performance.

Some underlying medical conditions, while occurring rarely, are known to cause much higher aPTT values. These medical conditions include lupus anticoagulants or deficiencies in the intrinsic (deficiency in factors IX or X) or extrinsic pathways (deficiency in factors VII) [30,31]. These conditions are not routinely checked for and are only diagnosed when advanced lab testing is ordered.

Established guidelines aim for a prolongation of aPTT by 1.5 to 2.5 times [11-13]. Since patients have different aPTT values before heparin is administered, the target value according to the guidelines is different. Furthermore, medical professionals may define individual anticoagulation targets that do not match a prolongation of 1.5 to 2.5 times the baseline value. Thus, we consider aPTT prediction to be a regression problem as Kong et al [16] and Smith et al [17] have done. A model that predicts aPTT several hours before blood is drawn and analyzed can serve as a valuable aid in adjusting the heparin dosing to meet the patient’s aPTT target earlier.

In principle, aPTT can be predicted continuously. However, to allow a comparison between models that make a single prediction based on measurements at a single point in time and a model that consumes the entire time series, we fixed 2 time points (at 12 hours and 24 hours after continuous treatment started). Models can use data available at 12 hours and make a prediction for 24 hours after continuous treatment starts. The cutoff after 12 hours is arbitrary and could be reasonably made at a different time. The second point in time is motivated by the observation that reaching the aPTT target within 24 hours is associated with favorable outcomes [6]. The recurrent neural network showed the best performance, and its predictions were analyzed in detail. Although most samples were predicted well, an unsolved problem is that rare cases exhibit a remarkably high aPTT and are not captured by the model. As mentioned earlier, underlying medical conditions are known to cause significantly longer aPTT. We hypothesized that, for significantly improved predictions, either testing of conditions that cause a long aPTT or much more frequent measurements of aPTT combined with dosing adjustments are required.

Recent literature on aPTT prediction after heparin treatment considers 3 distinct ranges [14,15,18]. In order to compare our model to those in the literature, we binned our predictions into subtherapeutic, therapeutic, and supratherapeutic as introduced by Ghassemi et al [14]. We observed that our model showed a higher recall and F1-score than did the other models. Arguably, the setup that we chose was the most difficult compared to the references since we predicted a single aPTT value 12 to 36 hours in advance. Others made predictions 4 to 6 hours [15] or 4 to 8 hours [14] in advance or averaged aPTT measurements between 4 and 24 hours [18].

Limitations

Other anticoagulants, such as warfarin or argatroban, were not considered. We expect that only a small sample of patients, if any, are receiving heparin together with anticoagulants and, therefore, decided not to take it into account as is common in similar studies [19].

It is well known that the laboratory conditions can affect the ranges of aPTT measurements [32]. The aPTT measurements were all reported by the same laboratory. Thus, the model may not be applicable to other centers and laboratories without parameter fine-tuning.

Modeling decisions that may negatively affect the model performance are the resampling of time series to 2-hour intervals. This resampling might miss significant changes in some variables. Furthermore, handling of missing data by forward and mean imputation could be improved by multiple imputation methods.

Conclusions

Anticoagulation therapy with heparin monitored by the aPTT laboratory assay is a widely used procedure in ICUs. It is well known that heparin dosing is challenging due to high interpatient variability. In the future, ML may help to suggest personalized dosing recommendations. We demonstrated that a model based on time series performs best.

Acknowledgments

We acknowledge financial support from the Open Access Publication Fund of Charité – Universitätsmedizin Berlin and the German Research Foundation (DFG).

Conflicts of Interest

FB reports grants from the German Federal Ministry of Education and Research, the German Federal Ministry of Health, the Berlin Institute of Health, the Hans Böckler Foundation, the Einstein Foundation, the Berlin University Alliance, and the Robert Koch Institute; and personal fees from Elsevier Publishing, Medtronic, and GE Healthcare outside the submitted work.
References


Abbreviations

- **APACHE II**: acute physiology and chronic health evaluation II
- **aPTT**: activated partial thromboplastin time
- **Charité**: Charité – Universitätsmedizin Berlin
- **ECMO**: extracorporeal membrane oxygenation
- **GRU**: gated recurrent unit
- **ICU**: intensive care unit
- **MIMIC**: Multiparameter Intelligent Monitoring in Intensive Care
- **ML**: machine learning
- **MSE**: mean-squared error
- **SAPS II**: simplified acute physiology score II
- **SOFA**: sequential organ failure assessment
- **SVR**: support vector machine regression
- **TISS-10**: therapeutic intervention scoring system

Edited by C Lovis, J Hefner; submitted 02.05.22; peer-reviewed by H Zhang, F Chen; comments to author 24.05.22; revised version received 17.07.22; accepted 11.08.22; published 13.10.22.

Please cite as:

Boie SD, Engelhardt LJ, Coenen N, Giesa N, Rubarth K, Menk M, Balzer F. A Recurrent Neural Network Model for Predicting Activated Partial Thromboplastin Time After Treatment With Heparin: Retrospective Study. JMIR Med Inform 2022;10(10):e39187 URL: https://medinform.jmir.org/2022/10/e39187 doi:10.2196/39187 PMID:36227653
©Sebastian Daniel Boie, Lilian Jo Engelhardt, Nicolas Coenen, Niklas Giesa, Kerstin Rubarth, Mario Menk, Felix Balzer. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 13.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.
Identifying Patients With Heart Failure Who Are Susceptible to De Novo Acute Kidney Injury: Machine Learning Approach

Caogen Hong\textsuperscript{1,2}, MSc; Zhoujian Sun\textsuperscript{3}, PhD; Yuzhe Hao\textsuperscript{2}, MSc; Zhanghuiya Dong\textsuperscript{2}, MSc; Zhaodan Gu\textsuperscript{2}, MSc; Zhengxing Huang\textsuperscript{4}, PhD

\textsuperscript{1}College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China
\textsuperscript{2}Jiangsu Automation Research Institute, Lianyungang, China
\textsuperscript{3}Research Center for Applied Mathematics and Machine Intelligence, Zhejiang Lab, Hangzhou, China
\textsuperscript{4}College of Computer Science and Technology, Zhejiang University, Hangzhou, China

Corresponding Author:
Zhoujian Sun, PhD
Research Center for Applied Mathematics and Machine Intelligence
Zhejiang Lab
Kechuang Ave, Zhongtai Subdistrict, Yuhang District
Hangzhou, 311121
China
Phone: 86 571 56390515
Fax: 86 518 85983716
Email: sunzj@zhejianglab.edu.cn

Abstract

Background: Studies have shown that more than half of patients with heart failure (HF) with acute kidney injury (AKI) have newonset AKI, and renal function evaluation markers such as estimated glomerular filtration rate are usually not repeatedly tested during the hospitalization. As an independent risk factor, delayed AKI recognition has been shown to be associated with the adverse events of patients with HF, such as chronic kidney disease and death.

Objective: The aim of this study is to develop and assess of an unsupervised machine learning model that identifies patients with HF and normal renal function but who are susceptible to de novo AKI.

Methods: We analyzed an electronic health record data set that included 5075 patients admitted for HF with normal renal function, from which 2 phenogroups were categorized using an unsupervised machine learning algorithm called K-means clustering. We then determined whether the inferred phenogroup index had the potential to be an essential risk indicator by conducting survival analysis, AKI prediction, and the hazard ratio test.

Results: The AKI incidence rate in the generated phenogroup 2 was significantly higher than that in phenogroup 1 (group 1: 106/2823, 3.75%; group 2: 259/2252, 11.50%; $P<.001$). The survival rate of phenogroup 2 was consistently lower than that of phenogroup 1 ($P<.005$). According to logistic regression, the univariate model using the phenogroup index achieved promising performance in AKI prediction (sensitivity 0.710). The generated phenogroup index was also significant in serving as a risk indicator for AKI (hazard ratio 3.20, 95% CI 2.55-4.01). Consistent results were yielded by applying the proposed model on an external validation data set extracted from Medical Information Mart for Intensive Care (MIMIC) III pertaining to 1006 patients with HF and normal renal function.

Conclusions: According to a machine learning analysis on electronic health record data, patients with HF who had normal renal function were clustered into separate phenogroups associated with different risk levels of de novo AKI. Our investigation suggests that using machine learning can facilitate patient phenogrouping and stratification in clinical settings where the identification of high-risk patients has been challenging.

(JMIR Med Inform 2022;10(10):e37484) doi:10.2196/37484

KEYWORDS
heart failure; acute kidney injury; unsupervised machine learning; risk stratification; phenogrouping
Introduction

Acute kidney injury (AKI) is a common disorder in patients with heart failure (HF), with the reported incidence rate varying from 7% to 38% in cardiology departments [1-3]. A recently conducted nationwide survey in China showed that about 85% of AKI incidents that occurred during cardiac hospitalization were ignored or were late to be identified [4,5]. As an independent risk factor, the delayed recognition of AKI has been proven to be associated with worse outcomes of patients with HF (eg, chronic kidney disease and mortality) [4,6]. To this end, the prompt identification of patients with HF at high-risk of AKI has great potential to improve clinical outcomes.

Although a few specific clinical markers (eg, estimated glomerular filtration rate [eGFR]) have been adopted to evaluate the renal function of patients with HF such that those at high risk of AKI can be identified, these markers lack the ability to screen de novo AKI patients who had normal renal function at admission [7,8]. Of note, several recently conducted population studies have indicated that more than half of the AKI that occurred in patients with HF were de novo [1-3]. To address this challenge, we attempted to clarify the characteristics of patients with HF who are susceptible to de novo AKI and developed a machine learning model for identification of HF patients with normal renal function but at high risk of de novo AKI.

As recently conducted cardiovascular studies have demonstrated that an unsupervised machine learning approach is able to model correlations among variables that contain prognostic information and cluster cohesive patients into 1 homogeneous phenogroup [9-11], we hypothesized that it can also be applied to identify patients with HF at high risk of de novo AKI. Recently, with the rapid development of hospital information systems, a large collection of electronic health records (EHRs) has become available that documents various types of patient information (eg, vital signs, laboratory test results) and treatments (eg, medication, surgery) and thus offers the considerable potential to implement a large-scale real-world analysis at a low expenditure. Therefore, in this study, we aimed to develop an EHR-based unsupervised machine learning analysis to group patients with HF and identify those who are susceptible to de novo AKI.

Methods

Study Population

The proposed retrospective study used a real-world data set obtained from the EHR system of the Chinese PLA General Hospital (PLAGH). The data set documented regular medical information in 84,705 hospitalizations of 29,699 patients who were diagnosed with HF in the PLAGH from 1998 to 2018. Adult patients with HF and normal renal function (eGFR >60 mL/min/1.73m$^2$ as calculated by the serum creatinine [SCr] version of the Chronic Kidney Disease Epidemiology Collaboration [CKD-EPI] equation [12] and without chronic kidney disease diagnosis) were considered for inclusion. Additionally, patients who did not have echocardiogram records were excluded. For patients with multiple hospitalizations, only the last hospitalization was reserved. The detailed preprocessing procedure is illustrated in Figure 1.
Ethics Approval
The study protocol was approved, with a waiver of consent granted on the basis of minimal harm and general impracticability by the health institutional review board of Zhejiang University (No. ZJU-2021-27).

Variable Selection and Machine Learning Model
In this study, 58 variables potentially associated with AKI, including demographics, vital sign measurements, medications, laboratories, operations, and echocardiogram exams, and routinely documented in EHRs at the admission stage of hospitalization were considered as candidates for analysis. To ensure that the most informative variables were selected and the correlation between variables could be diluted, we excluded variables with a missing rate larger than 30% or with a Pearson correlation coefficient >0.6 or that were documented fewer than 100 times in the raw EHR data set. As a result, 39 variables were included in the cohort. All continuous variables were transformed to standard normal distribution for the convenience of the unsupervised machine learning model (Table S1, Multimedia Appendix 1). Thereafter, we adopted multivariate imputation by chained equations [13] to impute the missing data.

We employed a simple yet effective unsupervised machine learning model called K-means clustering to categorize patients into different phenogroups [14]. The silhouette coefficient was applied to determine the optimal number of phenogroups [15]. We also adopted the nonlinear dimensionality reduction technique of t-distributed stochastic neighbor embedding [16] to visualize and evaluate the clustering results in a qualitative manner. The model was repeatedly run 1000 times to guarantee the achieved results stable.

Outcomes of Interest
The primary outcome was the incidence of AKI, which was defined according to the Kidney Disease: Improving Global Outcomes (KDIGO) standard [17], with the occurrence of AKI defined as the increase of SCr to ≥1.5 times the baseline in 7 days or the increase of SCr by ≥26.5 μmol/L within 48 hours. The secondary outcome was in-hospital mortality.

Characterization of Phenogroups
Once patients with HF were categorized into separate phenogroups, we measured the differences of variables in different groups. Continuous variables are reported as median and IQR (interquartile range). Categorical variables are reported as the frequencies and counts. Differences between groups were tested using the 1-way analysis of variance, Kruskal-Wallis test, or the chi-square test where appropriate. A P value of <.01 was considered statistically significant.

Discrimination of Phenogroups
We validated whether the phenogroup index generated by K-means clustering correlated with outcomes of interest by carrying out the following 3 experiments. First, Kaplan-Meier estimators with log-rank tests were conducted to analyze the time-to-event characteristics in different phenogroups. Second, we compared the prediction performance on AKI and in-hospital mortality to check whether the inferred phenogroup index was an effective risk predictor for outcomes of interest. Specifically, we selected the top-ranked 10 variables using a forward stepwise strategy with the Akaike information criterion and then developed 5 logistic regression (LR) models to predict the outcomes of interest. Model 1 used the phenogroup index as the univariate predictor. Model 2 used the top-ranked 10 variables as predictors. Model 3 used the top-ranked 10 variables and the phenogroup index. Model 4 used all 39 variables. Model 5 used all 39 variables and the phenogroup index. All models were trained by 70% of the data from the PLAGH data set and tested with the remaining 30% of data. Third, to evaluate whether the phenogroup index could achieve the competitive discriminative performance compared to the original variables with respect to the primary and secondary outcomes, we applied unadjusted Cox proportional hazard regression to examine hazard ratios (HRs), 95% CIs, and P values for all included original variables as well as the phenogroup index on both the whole PLAGH data set and the following subgroups: age (age <65 vs ≥65 years), sex, type of HF (acute vs chronic), diabetes mellitus, stroke, atrial fibrillation, coronary heart disease, anemia, and left ventricular ejection fraction (<40%, 40%-49%, and ≥50%). To assess continuous variables appropriately, we categorized all continuous variables in validation, and the cutoff points for these continuous variables are presented in online supplementary Table S2, Multimedia Appendix 1.

External Validation
We externally validated our model on a well-known open-source database, Medical Information Mart for Intensive Care (MIMIC)-III [18]. After a requisite preprocessing procedure (online supplementary, Figure S1), we prepared a MIMIC-III data set that contained 1006 patients with HF who had normal renal function. The model trained by the PLAGH data set was directly transferred onto the MIMIC-III data set. In detail, we compared the distance between the data of each patient in the MIMIC-III data set and the centroids of the derived phenogroups from the PLAGH data set and then assigned the patient into a phenogroup with the minimum Euclidean distance. After that, we assessed the survival rate and prediction performance of AKI and in-hospital mortality of the generated phenogroups from the MIMIC-III data set. As patients contained in the PLAGH data set were mainly from general wards in the PLAGH and patients included in the MIMIC-III data set were from intensive care units in the United States, there inevitably were statistical differences between the baseline characteristics of patients in the 2 data sets (Table S3, Multimedia Appendix 1). In this sense, the external validation was able to evaluate the stability of the proposed model in diverse clinical settings.

In this study, statistical and machine learning analysis was based on sklearn, lifelines, scipy package [19-21], and Python. We also report the centroids of the generated phenogroups from the PLAGH data set (Table S4, Multimedia Appendix 1), which may be nontrivial knowledge to assist clinicians in identifying their patients with HF at high risk of de novo AKI.
Results

Phenogroup Results

After preprocessing, 5075 hospitalizations and 39 variables (Table 1) were reserved for the PLAGH data set (median age 61 years, IQR 51-70 years; female 1723/5075, 32.39%; acute HF 1723/5075, 33.95%). Using K-means clustering, we naturally separated patients into 2 basically nonoverlapping phenogroups, where the number of clusters was suggested by the silhouette coefficient test (Figure S1, Multimedia Appendix 1). Similar results were found using t-distributed stochastic neighbor embedding visualization (Figure S2, Multimedia Appendix 1).

Table 1. Included variables for clustering.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td>Age, sex</td>
</tr>
<tr>
<td>Disease</td>
<td>Acute/chronic HF, atrial fibrillation, cardiomyopathy, coronary heart disease, diabetes, stroke, valvular heart disease</td>
</tr>
<tr>
<td>Medication</td>
<td>Angiotensin-converting enzyme inhibitor/angiotensin receptor blocker, anticoagulant, antiplatelet, beta blocker, calcium channel blocker, diuretic, positive inotropic drug, vasodilator</td>
</tr>
<tr>
<td>Echocardiography</td>
<td>Left ventricular ejection function</td>
</tr>
<tr>
<td>Laboratory result</td>
<td>Alanine aminotransferase, aspartate transaminase, estimated glomerular filtration rate, gamma-glutamyl transferase, hemoglobin, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, N-terminal probrain natriuretic peptide, serum calcium, serum potassium, serum sodium, serum urea, total bilirubin, total serum protein, triglyceride, troponin T</td>
</tr>
<tr>
<td>Operation</td>
<td>Angiography percutaneous coronary intervention</td>
</tr>
<tr>
<td>Vital sign</td>
<td>BMI, diastolic blood pressure, systolic blood pressure</td>
</tr>
</tbody>
</table>

*Only drugs used in the first 48 hours after admission were included to ensure the drug usage could reflect the patient admission status.

Characteristics of Phenogroups

Table 2 illustrates the baseline characteristics of the PLAGH data set and the 2 derived phenogroups. Compared to phenogroup 1, phenogroup 2 had a higher rates of AKI (group 1: 106/2823, 3.75%; group 2: 259/2252, 11.50%; P<.001) and in-hospital mortality (phenogroup 1: 21/2823, 0.74%; phenogroup 2: 118/2252, 5.24%; P<.001). In addition, patients in phenogroup 2 were generally older than those in phenogroup 1 (58 vs 65 years; P<.001).

As can be seen in Table 2, there are more patients diagnosed with acute HF in phenogroup 2 than those in phenogroup 1 (phenogroup 1: 738/2823, 26.14%; phenogroup 2: 985/2252, 43.74%; P<.001). Moreover, cardiac function of patients in phenogroup 2 was worse than that in phenogroup 1. Specifically, there were statistical differences between patients in phenogroup 1 and phenogroup 2 in terms of left ventricular ejection fraction (50% vs 41%; P<.001), diastolic blood pressure (77 mmHg vs 70 mmHg; P<.001), systolic blood pressure (130 mmHg vs 118 mmHg; P<.001), N-terminal pro-brain natriuretic peptide (572 pg/mL vs 2680 pg/mL; P<.001), hemoglobin (143 g/L vs 129 g/L; P<.001), atrial fibrillation (phenogroup 1: 526/2823, 18.63%; phenogroup 2: 595/2252, 26.42%; P<.001), diuretic usage (phenogroup 1: 1608/2823, 56.96%; phenogroup 2: 1799/2252, 79.88%; P<.001), and positive inotropic drug usage (phenogroup 1: 778/2823, 27.56%; phenogroup 2: 1089/2252, 48.36%; P<.001). Furthermore, phenogroup 2 had higher troponin T levels (0.01 ng/mL vs 0.02 ng/mL; P<.001), indicating that there were more patients in phenogroup 2 who underwent myocardial damage. Patients in phenogroup 2 had higher values of gamma-glutamyl transferase (31.70 IU/L vs 40.30 IU/L; P<.001), total bilirubin (12.79 μmol/L vs 15.85 μmol/L; P<.001), and aspartate aminotransferase (19.60 IU/L vs 24.29 IU/L; P<.001), indicating that patients in phenogroup 2 might have worse liver function compared with phenogroup 1. Moreover, although we had excluded patients with renal dysfunction in advance, patients in phenogroup 2 had worse eGFR values (92.06 mL/min/1.73 m² vs 81.85 mL/min/1.73 m²; P<.001) and urea (5.46 mmol/L vs mmol/L; P<.001). These findings demonstrated that patients in phenogroup 2 had relatively worse kidney function. Furthermore, patients in phenogroup 2 used less angiotensin-converting enzyme inhibitor/angiotensin receptor blocker (phenogroup 1: 1531/2823, 54.23%; phenogroup 2: 1016/2252, 45.11%; P<.001), calcium channel blocker (phenogroup 1: 789/2252, 37.95%; phenogroup 2: 321/2252, 15.25%; P<.001), and antiplatelets (phenogroup 1: 1914/2823, 67.80%; phenogroup 2: 1384/2823, 61.45%; P<.001). It was worth noting that patients in phenogroup 2 had higher lipid levels (low-density lipoprotein cholesterol and triglyceride) and BMI (25.88 kg/m² vs 23.05 kg/m²; P<.001). Compared to phenogroup 1, phenogroup 2 also received less angiography (phenogroup 1: 1311/2823, 46.44%; phenogroup 2: 1311/2252, 30.95%; P<.001) and percutaneous coronary intervention (phenogroup 1: 640/2823, 21.96%; phenogroup 2: 349/2252, 15.50%; P<.001). Comprehensive baseline characteristics including all 58 candidate variables are listed in Table S5, Multimedia Appendix 1.
Table 2. Baseline characteristics of the PLA General Hospital data set and the generated phenogroups.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Population (N=5075)</th>
<th>Phenogroup 1 (n=2823)</th>
<th>Phenogroup 2 (n=2252)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Feature of interest, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKI&lt;sup&gt;a&lt;/sup&gt;</td>
<td>365 (7.19)</td>
<td>106 (3.75)</td>
<td>259 (11.50)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>In-hospital mortality</td>
<td>139 (2.74)</td>
<td>21 (0.74)</td>
<td>118 (5.24)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Demographic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>61 (51-70)</td>
<td>58 (48-67)</td>
<td>65 (55-75)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), median (IQR)</td>
<td>24.60 (22.46-27.08)</td>
<td>25.88 (23.87-28.08)</td>
<td>23.05 (20.95-25.01)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DBP&lt;sup&gt;b&lt;/sup&gt; (mmHg), median (IQR)</td>
<td>74 (67-81)</td>
<td>77 (70-85)</td>
<td>70 (64-78)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SBP&lt;sup&gt;c&lt;/sup&gt; (mmHg), median (IQR)</td>
<td>125 (113-138)</td>
<td>130 (119-143)</td>
<td>118 (106-130)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>3431 (67.61)</td>
<td>1943 (68.83)</td>
<td>1488 (66.07)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Disease, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute HF</td>
<td>1723 (33.95)</td>
<td>738 (26.14)</td>
<td>985 (43.73%)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chronic HF</td>
<td>3352 (66.05)</td>
<td>2075 (73.86)</td>
<td>1267 (56.26%)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>AF&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1121 (22.09)</td>
<td>526 (18.63)</td>
<td>595 (26.42)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Cardiomyopathy</td>
<td>941 (18.54)</td>
<td>497 (17.61)</td>
<td>444 (19.71)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CHD&lt;sup&gt;f&lt;/sup&gt;</td>
<td>2928 (57.69)</td>
<td>1660 (58.80)</td>
<td>1268 (56.30)</td>
<td>.07</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2002 (39.44)</td>
<td>1041 (36.88)</td>
<td>961 (42.67)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stroke</td>
<td>485 (9.56)</td>
<td>282 (9.99)</td>
<td>233 (10.35)</td>
<td>.09</td>
</tr>
<tr>
<td>VHD&lt;sup&gt;g&lt;/sup&gt;</td>
<td>616 (12.13)</td>
<td>336 (11.90)</td>
<td>280 (12.43)</td>
<td>.57</td>
</tr>
<tr>
<td><strong>Medication, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACEI/ARB&lt;sup&gt;h&lt;/sup&gt;</td>
<td>2547 (50.18)</td>
<td>1531 (54.23)</td>
<td>1016 (45.11)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Anticoagulant</td>
<td>1927 (37.97)</td>
<td>989 (35.03)</td>
<td>938 (41.65)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Antiplatelet</td>
<td>3298 (64.99)</td>
<td>1914 (67.80)</td>
<td>1384 (61.45)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Beta blocker</td>
<td>3428 (67.54)</td>
<td>1981 (70.17)</td>
<td>1447 (64.25)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CCB&lt;sup&gt;i&lt;/sup&gt;</td>
<td>1110 (21.87)</td>
<td>789 (27.95)</td>
<td>321 (14.25)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Diuretic</td>
<td>3407 (67.13)</td>
<td>1608 (56.96)</td>
<td>1799 (79.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Positive inotropic drugs</td>
<td>1867 (36.79)</td>
<td>778 (27.56)</td>
<td>1089 (48.36)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vasodilator</td>
<td>3103 (61.14)</td>
<td>1698 (60.15)</td>
<td>1405 (62.39)</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Echocardiogram</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVEF&lt;sup&gt;j&lt;/sup&gt;, median (IQR)</td>
<td>46 (35-56)</td>
<td>50 (39-58)</td>
<td>41 (31-54)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;40%, n (%)</td>
<td>1716 (33.81)</td>
<td>719 (25.47)</td>
<td>997 (44.27)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>40%-50%, n (%)</td>
<td>1174 (23.13)</td>
<td>690 (24.44)</td>
<td>484 (21.49)</td>
<td>.05</td>
</tr>
<tr>
<td>≥50%, n (%)</td>
<td>2185 (42.86)</td>
<td>1414 (50.09)</td>
<td>771 (34.24)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Laboratory result, median (IQR)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALT&lt;sup&gt;k&lt;/sup&gt;, (IU/L)</td>
<td>21.00 (14.39-33.79)</td>
<td>20.80 (14.70-31.99)</td>
<td>21.54 (13.80-36.49)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>AST&lt;sup&lt;l&lt;/sup&gt;, (IU/L)</td>
<td>21.29 (16.29-30.50)</td>
<td>19.60 (15.50-26.00)</td>
<td>24.29 (18.09-38.80)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Calcium (mmol/L)</td>
<td>2.24 (2.16-2.33)</td>
<td>2.28 (2.21-2.36)</td>
<td>2.19 (2.10-2.27)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>eGFR&lt;sup&gt;m&lt;/sup&gt; (mL/min/1.73 m&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>87.62 (75.65-98.80)</td>
<td>92.06 (80.84-101.91)</td>
<td>81.85 (70.90-92.91)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>GGT&lt;sup&gt;n&lt;/sup&gt; (IU/L)</td>
<td>34.80 (21.90-63.79)</td>
<td>31.70 (21.30-54.89)</td>
<td>40.30 (23.09-75.00)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
### Survival Analysis

As the prevalence of AKI and in-hospital mortality had a significant difference between the generated phenogroups, phenogroup 1 was intuitively labeled as “low-risk” and phenogroup 2 as “high-risk.” We further investigated whether the generated phenogroup index could serve as an essential risk indicator for clinical outcomes of interest.

Figure 2 shows the survival difference with respect to AKI and in-hospital mortality between the generated “high-risk” and “low-risk” phenogroups from both the PLAGH data set and the external validation MIMIC-III data set. For AKI, the curves of phenogroup 2 were lower than the curves of phenogroup 1 in both development and external validation data sets (PLAGH: \( P = .004 \); MIMIC-III: \( P = .002 \)). In addition, we found that most AKI events often happened in the first few days of hospitalization in both the PLAGH and MIMIC-III data sets. This finding was in line with the literature [7,8]. For in-hospital mortality, the curves of phenogroup 2 were consistently lower than the curves of phenogroup 1 (PLAGH: \( P = .002 \); MIMIC-III: \( P = .01 \)). In consideration of the baseline difference between the PLAGH data set and MIMIC-III data set, the results demonstrated that our model was robust in discriminating between high-risk and low-risk patients and easily transferable to different clinical settings.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Population (N=5075)</th>
<th>Phenogroup 1 (n=2823)</th>
<th>Phenogroup 2 (n=2252)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDL-C(^b) (mmol/L)</td>
<td>1.02 (0.85-1.22)</td>
<td>1.04 (0.88-1.22)</td>
<td>1.01 (0.82-1.22)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hemoglobin, g/L</td>
<td>137 (124-150)</td>
<td>143 (132-154)</td>
<td>129 (116-142)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LDL-C(^p) (mmol/L)</td>
<td>2.25 (1.79-2.81)</td>
<td>2.46 (1.96-3.05)</td>
<td>2.04 (1.62-2.48)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>NT-pro-BNP(^q) (pg/mL)</td>
<td>1216 (422-2950)</td>
<td>572 (225-1319)</td>
<td>2680 (1355-5188)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Potassium (mmol/L)</td>
<td>3.89 (3.62-4.17)</td>
<td>3.87 (3.62-4.13)</td>
<td>3.91 (3.61-4.20)</td>
<td>.005</td>
</tr>
<tr>
<td>Sodium (mmol/L)</td>
<td>140.70 (138.10-142.70)</td>
<td>141.30 (139.40-143.20)</td>
<td>139.40 (136.30-142.00)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Total bilirubin (( \mu )mol/L)</td>
<td>13.69 (9.80-19.90)</td>
<td>12.79 (9.40-17.40)</td>
<td>15.85 (10.39-24.60)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Total protein (g/L)</td>
<td>67.5 (63.3-71.8)</td>
<td>69.2 (65.8-73.3)</td>
<td>65.1 (60.4-69.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Triglyceride (mmol/L)</td>
<td>1.11 (0.82-1.59)</td>
<td>1.34 (0.98-1.87)</td>
<td>0.92 (0.72-1.21)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Troponin T (ng/mL)</td>
<td>0.01 (0.01-0.04)</td>
<td>0.01 (0.00-0.02)</td>
<td>0.02 (0.01-0.10)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Urea (mmol/L)</td>
<td>5.84 (4.73-7.25)</td>
<td>5.46 (4.51-6.60)</td>
<td>6.45 (5.11-8.12)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

### Operation, n (%)

- **Angiography**: 2008 (29.57) vs. 1311 (46.44) vs. 697 (30.95), \( P = .001 \)
- **PCI\(^r\)**: 969 (19.09) vs. 620 (21.96) vs. 349 (15.50), \( P = .001 \)

\(^a\)AKI: acuted kidney injury.  
\(^b\)DBP: diastolic blood pressure.  
\(^c\)SBP: systolic blood pressure.  
\(^d\)HF: heart failure.  
\(^e\)AF: atrial fibrillation.  
\(^f\)CHD: coronary artery disease.  
\(^g\)VHD: valvular heart disease.  
\(^h\)ACEI/ARB: angiotensin-converting enzyme inhibitor/angiotensin receptor blocker.  
\(^i\)CCB: calcium channel blocker.  
\(^j\)LVEF: left ventricular ejection fraction.  
\(^k\)ALT: alanine aminotransferase.  
\(^l\)AST: aspartate transaminase.  
\(^m\)eGFR: estimated glomerular filtration rate.  
\(^n\)GGT: gamma-glutamyl transferase.  
\(^o\)HDL-C: high-density lipoprotein cholesterol.  
\(^p\)LDL-C: low-density lipoprotein cholesterol.  
\(^q\)NT-pro-BNP: N-terminal probrain natriuretic peptide.  
\(^r\)PCI: percutaneous coronary intervention.
Figure 2. Kaplan-Meier curves for AKI and in-hospital mortality in the development (PLAGH) and external validation (MIMIC-III) data sets. AKI: acute kidney injury; MIMIC: Medical Information Mart for Intensive Care; PLAGH: PLA General Hospital.

Outcome Prediction
Table 3 compares the prediction performances of the 5 LR models. Sensitivity, specificity, and concordance statistics are reported for the prediction performance evaluation. As the false-negative prediction (ie, neglecting AKI) may lead to extremely negative consequences, we mainly compared the sensitivity performance among the 5 models. The threshold of sensitivity and specificity was 0.5 in all experiments, and the selected top-10 variables are listed in Table S6, Multimedia Appendix 1. The results showed that the phenogroup index was an essential risk predictor of outcomes. For one, Model 1 used 1 variable (the phenogroup index) as the predictor and achieved promising sensitivity in terms of AKI (0.710) and in-hospital mortality (0.820) among the 5 prediction models with the PLAGH data set. For another, the prediction performance of Model 1 remained quite stable in the external validation (AKI sensitivity 0.760; in-hospital mortality sensitivity 0.826), while there existed significant degradation of performance in the other prediction models.
Table 3. Prediction performance comparison.

<table>
<thead>
<tr>
<th>Model by task</th>
<th>PLAGH(^a) data set (development)</th>
<th>MIMIC-III(^b) data set (validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>AKI(^d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>0.710</td>
<td>0.577</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.647</td>
<td>0.638</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.679</td>
<td>0.723</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.737</td>
<td>0.753</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.718</td>
<td>0.746</td>
</tr>
<tr>
<td>In-hospital mortality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>0.849</td>
<td>0.568</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.791</td>
<td>0.736</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.820</td>
<td>0.763</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.835</td>
<td>0.809</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.856</td>
<td>0.812</td>
</tr>
</tbody>
</table>

\(^a\)PLAGH: PLA General Hospital.
\(^b\)MIMIC-III: Medical Information Mart for Intensive Care III.
\(^c\)C-statistic: concordance statistic.
\(^d\)AKI: acute kidney injury.

**HR Comparison**

We used unadjusted Cox proportional hazard regression to determine whether the phenogroup index can act as an essential risk stratification indicator in comparison with the original 39 included variables. The top-ranked 10 variables with the highest HR are listed in Figure 3 (full list is available from Figure S3, Multimedia Appendix 1). The results showed that the HR of the phenogroup index was ranked second in AKI analysis and first in in-hospital mortality analysis, indicating that the phenogroup index can be an effective risk stratification indicator compared with the original variables. Of further note, although troponin T was ranked first for AKI analysis, it was not appropriate for univariate risk indicators since only 16.73% (849/5075) of patients in the PLAGH data set had abnormal records in troponin T. Using troponin T as the indicator only achieved a sensitivity of 0.431, which was significantly lower than the performance of the phenogroup index (0.710). The association between the generated phenogroup index and risk of AKI (in-hospital mortality) was consistent in all examined subgroups (Figure 4).

**Figure 3.** Hazard ratios of top-ranked 10 discriminative features for (a) acute kidney injury and (b) in-hospital mortality from the PLA General Hospital data set. AST: aspartate aminotransferase; eGFR: estimated glomerular filtration rate; NT-pro-BNP: N-terminal probrain natriuretic peptide. *Anemia was defined as hemoglobin <135 g/L for men and hemoglobin <120 g/L for women. All units of variables in this figure are same as the units in Table 2.
**Discussion**

**Principal Findings**

We explored the potential of using a large volume of EHR data to cluster patients with HF and identify those with normal renal function but susceptible to de novo AKI via an unsupervised machine learning model. The experimental results showed that there was significant difference in AKI and in-hospital mortality occurrence between the 2 phenogroups generated from EHR data. As EHR is a real-world, readily available data source containing rich medical information of thousands of patients, our study demonstrated that it was possible for researchers to answer important clinical and scientific questions effectively by exploiting the huge potential of EHR data via machine learning techniques at a fraction of the resource cost that would have been required using traditional approaches [22,23].

We demonstrated that HF patients with normal renal function can be naturally separated into a “high-risk phenogroup,” of patients susceptible to de novo AKI and a “low-risk phenogroup” who were not. Patients in high-risk phenogroup were typically older, more susceptible to multi-organ dysfunction and anemia, and had significantly higher in-hospital mortality than did those in the low-risk phenogroup. These findings were in line with recent studies [17,24] and warrant further assessment. We found that patients in the high-risk phenogroup had lower levels of lipid and BMI than did those in the low-risk group. These findings are consistent with previous studies reporting that worse cardiac function may cause malnutrition [25] and a decrease of lipid level [26]. Of note, worse cardiac function was also associated with hemodynamic instability, which influences the choice of oral medication strategies [27]. We observed that patients in the high-risk phenogroup received less medication (angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, calcium channel blocker, and beta blockers) than did those in the low-risk phenogroup. On the contrary, we found that patients in low-risk phenogroup were likely to receive percutaneous coronary intervention (PCI) during their stay at the emergency care unit or in hospitalization to revascularize the stable hemodynamic level such that the perfusion of the kidney could be improved and the risk of AKI significantly alleviated. This finding is consistent with previous findings, emphasizing the benefit of timely revascularization [28].

Identification of patients with HF with normal renal function but at high-risk of de novo AKI is a major challenge in HF treatment management. Clinicians have highlighted the need for more effective methods to perform this important clinical task [29]. In this study, we illustrated that machine learning analysis can tackle this challenge by providing deep integration of the comprehensive clinical variables routinely documented in EHR data. As observed in the present study, the phenogroup index generated by an unsupervised machine learning approach, as a latent representation of 39 original variables and their interactions, exhibited a sensitivity of 0.710 and 0.760 on the development data set (PLAGH) and the external validation data set (MIMIC-III). In this sense, the generated phenogroups from raw EHR data are meaningful and can be translated into actionable information for clinical decision-making. On the
contrary, all other LR models met a serious overfitting problem
due to the fact that the included variables had different
 distributions between the development (FLAGH) and external
validation (MIMIC-III) data sets (as can be seen in Table S3.
Multimedia Appendix 1). Inevitably, this issue caused a
significant performance degeneration in the external validation.
In consideration of the baseline difference between the FLAGH
data set and the MIMIC-III data set, the results suggested that
the generated phenogroup index was able to act as an essential
de novo AKI risk indicator for patients with HF and normal
renal function and can be smoothly applied in different clinical
settings and in different patient populations. In fact, machine
learning algorithms can handle a large volume of variables and
a vast number of variable-variable interactions in each patient.
This merit effectively individualizes risk assessment and
remedies many of the limitations of standard statistical models
[22].

Our study has potentially important clinical ramifications. For
one, as AKI risk is often underestimated or neglected in patient
with HF, especially those with normal renal function [5], our
study provided a new perspective for identifying patients with
HF and normal renal function but who are at high risk of AKI.
For another, in comparison with recent studies that focused on
finding new biomarkers for AKI prediction or detection [30],
we adopted an improved alternative strategy that used machine
learning techniques to explore readily available clinical data to
identify patients with HF at high risk of de novo AKI. Such
meaningful use of EHR data may provide the best available
evidence to assist clinical decision-making. It should be noted
that these improvements may be enhanced by mining a large
volume of readily available EHR data, which in turn may
provide a new avenue for improving any given machine learning
algorithm.

Limitations
Several limitations of this study should be acknowledged. First,
this is a single-institution study. Although we have evaluated
our model on an external validation data set extracted from
MIMIC-III, the methods perform less well in other situations due to the lack of sufficient external validation
samples collected from different medical facilities and in different clinical settings. Second, our study was limited by its
retrospective design, and all analyses were purely observational.
Although we found that there were distinct variables associated
with increased risks of de novo AKI and in-hospital mortality,
these nonrandomized comparisons should be interpreted
cautiously in this context, and the prognostic ability of our model
needs to be supported by validation in prospective studies. Third,
considering the sensitivity and the specificity for AKI
forecasting, our model was relatively sensitive but not very
specific. Despite the influence of false-positive classification
being limited in this study, further study will be required to
enable machine learning–based analysis to capture the salient
features distinguishing high- from low-risk cases, such that the
prediction performance of our model can be improved.

Conclusions
This study demonstrated that unsupervised machine
learning–based EHR analysis is able to separate patients with
HF and normal renal function into mutually exclusive
phenogroups that correspond to saliently distinct AKI risk levels.
Our investigation paves the way for developing an easy-to-use,
broadly available model that allows the identification of patients
with HF at high-risk of de novo AKI and may help improve
outcomes in HF, offering a crucial advantage over traditional
techniques for patient phenogrouping and clinical risk
stratification.

Acknowledgments
The contribution of investigators and clinical coordinators are duly acknowledged.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Experimental data set introduction and detailed experiment results.
[DOC File, 2618 KB - medinform_v10i10e37484_app1.doc ]

References
definitions (RIFLE, AKIN, and KDIGO) of acute kidney injury for the prediction of outcomes in acute decompensated
ESC Heart Fail 2020 Apr;7(2):415-422 [FREE Full text] [doi: 10.1002/ehf2.12595] [Medline: 32059081]


**Abbreviations**

- **AF**: atrial fibrillation
- **AKI**: acute kidney injury
- **CHD**: coronary heart disease
- **CKD**: chronic kidney disease
- **CKD-EPI**: Chronic Kidney Disease Epidemiology Collaboration
- **eGFR**: estimated glomerular filtration rate
- **EHR**: electronic health record
- **HF**: heart failure
- **HR**: hazard ratio
- **KDIGO**: Kidney Disease: Improving Global Outcomes
- **LR**: logistic regression
- **MIMIC**: Medical Information Mart for Intensive Care
- **PLAGH**: Chinese PLA General Hospital
- **SCr**: serum creatinine

©Caogen Hong, Zhoujian Sun, Yuzhe Hao, Zhanghuiya Dong, Zhaodan Gu, Zhengxing Huang. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 14.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.